



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

Conceptualising research quality in medicine for evaluative bibliometrics

Andersen, Jens Peter

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Andersen, J. P. (2013). *Conceptualising research quality in medicine for evaluative bibliometrics*. Det Humanistiske Fakultet, Københavns Universitet.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.



PhD thesis
Jens Peter Andersen

Conceptualising research quality in medicine for evaluative bibliometrics

Academic advisor: Jesper Wiborg Schneider

Submitted: 26/07/2013

Conceptualising research quality in medicine for evaluative bibliometrics

Jens Peter Andersen

PhD thesis from
Research Programme for Information Studies
Faculty of Humanities, University of Copenhagen

CIP - Cataloging in publication

Andersen, Jens Peter

Conceptualising research quality in medicine for evaluative bibliometrics / Jens Peter Andersen- Aalborg: Research Programme for Information Studies, Faculty of Humanities, Copenhagen University, Denmark, 2013. x. 398 p.

<http://iva.ku.dk/jenspeterandersenthesis>

ISBN: 9788774153283

Konceptualisering af forskningskvalitet indenfor medicin

Jens Peter Andersen

Ph.d.-afhandling fra
Forskningsprogram for informationsstudier
Det Humanistiske Fakultet, Københavns Universitet

For Linnea and Anette

Acknowledgments

A great number of people have helped in the creation of this dissertation and I am thankful to each and every one of them. Without their help and support, this dissertation and the process of writing it would not have been the same.

My first and foremost thank goes to my advisor and mentor, Jesper W. Schneider. You have played a major role in shaping this dissertation, and your subtle way of guiding me in the right direction has allowed me to grow with the task. This is not only true for my work with the dissertation, but your trust and guidance has also helped me in the many other areas that are part of becoming a researcher. Before even starting on this path, a number of people inspired and supported me to go in this direction at all: Pia Borlund offered her help from the very beginning, and together with Conni Skrubbeltrang and Hans Gregersen was invaluable in creating the initial proposals and a platform for the actual PhD position. Also Birgitta Olander and Fredrik Åström offered their council and trust in those initial stages.

I would like to thank the NORSLIS research school and the ISSI doctoral forums for their fruitful PhD workshops and forums, and in particular the significant feedback from Dick Klavans, Paul Wouters and Peter Ingwersen. In addition to the curriculum, I have met many exciting people through both institutions. Particularly Björn Hammarfelt has been a good friend and collaborator; I hope we will have more opportunities to work together in the future.

I would also like to thank the many people I have met at the ISSI and STI conferences during my time as a PhD student. You have been most welcoming and it is always interesting to discuss bibliometrics with you. Especially I would like to thank Ludo Waltman, Thed van Leeuwen, Rodrigo Costas, Dag Aksnes, Carolin Michels, Pei-Shan Chi, Truyken Ossenblock, Stefanie Haustein and Grant Lewison for their advice, discussions and company.

A key element of this dissertation is the input from medical researchers in both interview sessions and an online survey. All the participants in these studies have been incredibly important for the study, and the dialogues during the interviews were inspiring and eye-opening. I am humbled by the interest and time these participants have spent on my project.

My colleagues at the medical library at Aalborg University Hospital have all been very understanding and supportive. Conni, Hanne, Jakob, Jette, Marianne, Louise, Pernille, Kristin and Tenna - you make me look forward to going to work every day. Also my other colleagues at Aalborg Hospital Science and Innovation Center have been incredibly supportive and I appreciate the advice and help I have received from all of you.

My final appreciation goes to my partner in life, the universe and everything; Anette. I could not imagine a better friend, wife and lover than you - you have been there for me and our daughter, challenged me, made me want to try harder, aspire for more and never give up.

Abstract

The use of the term “research quality” in bibliometric research assessment is a problematic, yet common, practice. While the concept can be operationalised as a matter of successfully executing and publishing research (e.g. Smart, 2005; Andras, 2011), there are several other elements of research which could be considered qualities. Such elements could be e.g. the practical implications of research, the effects on society or the adaptation by other research areas. Traditional evaluative bibliometrics often measure the impact of research, which is seen as an aspect of research quality. However, impact can also be interpreted in different ways, but yet there is a semantic relationship between the interpretation of research quality and impact; a relationship which is most likely an overlap between certain aspects of research quality and bibliometric quantities. The aim of this dissertation was to conceptualise research quality, through a description of the qualities of the concept relating to the formalised dissemination of research, and describing the dimensions of the concept. The purpose of this conceptualisation was an articulation of the interface between research quality and evaluative bibliometrics. The study is delimited to the medical field; a field with high productivity, an established scientific society and internal perceptions (formal as well as informal) of research quality.

A combination of qualitative and quantitative methods were used to investigate the concept of research quality. In the initial phase, interviews were used to obtain qualitative statements on how research quality is perceived by researchers and practitioners in the medical field. 14 people from Aalborg University Hospital participated in these interviews, and codified statements were extracted from transcripts of the interviews. These were validated through an online survey sent to medical researchers from most of Europe, North America and Australia. A total of 279 complete responses to the survey were collected, and factor analysis was used to analyse the underlying structure of the included variables. Based on this, important variables, factors and interactions were identified. The factors were further qualified by relating them back to the qualitative data they were originally derived from. The resulting interlinked, narrated factors were used to create two models of research quality, one in a research process context, and one highlighting the descriptions of dimensions and quantitations of research quality. In both models, the central elements are three dimensions of research quality; dissemination, policy effects and health effects. Each of these dimensions, or impact types, provide a partial answer to definition of the quality of research.

The results of the present dissertation contribute to the progression of bibliometric research through an elucidation of the research quality concept and its interface with bibliometric methodology. It also adds to the debate about bibliometric terminology by elaborating the impact aspect of research quality. Finally, it calls for further research on different document types, as well as citation context and path studies.

Abstract in danish

Brugen af ordet "forskningskvalitet" indenfor bibliometrisk forskningsevaluering er en problematisk, men alligevel almindelig, praksis. Omend begrebet kan operationalises som den succesfulde udførsel og publicering af forskning (f.eks. Smart, 2005; Andras, 2011), er der adskillige andre elementer af forskning der kan anses som kvaliteter. Sådanne elementer kunne f.eks. være forskningens praktiske betydning, den samfundsmæssige betydning og andre forskningsområders tilegnelse. Traditionel, evaluerende bibliometri måler ofte hvad der betegnes som forskningens "impact", der ses som et aspekt af forskningskvalitet. Impact kan dog også fortolkes forskelligt, men der er alligevel en semantisk relation imellem fortolkningen af forskningskvalitet og impact. Denne relation kan sandsynligvis betegnes som et overlap imellem visse aspekter af forskningskvalitet og bibliometriske mål. Formålet med denne afhandling var at konceptualisere forskningskvalitet, igennem en beskrivelse af de af begrebets kvaliteter, der er relateret til den formaliserede udbredning af forskning, og en beskrivelse af begrebets dimensioner. Formålet med denne konceptualisering var en italesættelse af berøringsfladen imellem forskningskvalitet og bibliometri. Undersøgelsen er afgrænset til det medicinske område; et område præget af høj produktivitet, et etableret videnskabeligt samfund og en intern opfattelse (formel som uformel) af forskningskvalitet.

Der blev anvendt en kombination af kvalitative og kvantitative metoder til at undersøge begrebet forskningskvalitet. I den indeledende fase blev interviews brugt til at indsamle kvalitative udtalelser om, hvordan forskningskvalitet bliver opfattet af forskere og praktikere i det medicinske område. 14 mennesker fra Aalborg Universitetshospital deltog i disse interviews, og kodede udtalelser blev udtrukket fra interviewtranskriberingerne. Disse blev efterfølgende valideret igennem et online spørgeskema, sendt til medicinske forskere fra størstedelen af Europa, Nordamerika og Australien. Ialt blev der indsamlet 279 fuldstændige svar fra spørgeskemaet, og der blev udført faktoranalyse for at undersøge den underliggende struktur i de inkluderede variable. Med udgangspunkt i dette identificeredes afgørende variable, faktorer og interaktioner herimellem. Faktorerne blev yderligere kvalificeret ved at holde dem op imod de kvalitative data, de oprindeligt blev afledt af. De endelige, relaterede, beskrevne faktorer blev brugt til at fremstille to modeller over forskningskvalitet. Den ene model sætter forskningskvalitet i en forskningsprocesskontekst, imens den anden belyser beskrivelsen af forskningskvalitetsbegrebets dimensioner og kvantiteringer. I begge modeller er de centrale elementer tre dimensioner af forskningskvalitet; udbredelse (disseminering), politikeffekter og helbredseffekter. Hver af disse dimensioner, eller impact typer, bidrager med en del af svaret til en definering af forskningskvalitet.

Resultaterne af den forhåndenværende afhandling bidrager til progressionen af bibliometrisk forskning igennem en præcisering af forskningskvalitetsbegrebet og dets berøringsflade med bibliometrisk metodologi. Afhandlingen bidrager også til debatten om bibliometrisk terminologi ved at udbygge impactaspektet af forskningskvalitet. Endelig lægger det op til yderligere forskning indenfor forskellige dokumenttyper, såvel som citationskontekst- og "citation path"-undersøgelser.

Contents

1	Introduction	11
1.1	Objectives of the dissertation	13
1.2	Research questions	15
1.3	Structure of the dissertation	16
2	Research quality	17
2.1	The research quality concept	17
2.2	Research quality and impact assessment	19
2.2.1	<i>Negative findings, fraud and retractions</i>	20
2.2.2	<i>Peer assessment</i>	21
2.3	Summary	30
3	Measuring research quality	33
3.1	Quantitation & measurement	34
3.1.1	<i>Quantitation of quality</i>	35
3.1.2	<i>Measurement & metrics</i>	37
3.2	Evaluative bibliometrics	39
3.2.1	<i>Bibliometric foundation</i>	40
3.2.2	<i>Bibliometric units</i>	41
3.2.3	<i>Citation analysis</i>	44
3.2.4	<i>Impact</i>	48
3.3	Referencing and citation theory	50
4	Research quality in medicine	55
4.1	A very brief history of medicine	57
4.2	Evidence-based medicine and scientific communication	59
5	Positioning the dissertation	65

6	Methods and materials	67
6.1	Study Design	68
6.2	Interview study	69
6.2.1	<i>Pilot interview study</i>	74
6.2.2	<i>Script revision</i>	75
6.2.3	<i>Main interview study</i>	75
6.3	Online survey	79
6.4	Factor analysis	85
6.5	Demographic analysis	88
6.5.1	<i>Demographic variables summarised</i>	89
6.6	Creating narratives	94
7	Results	95
7.1	Data analysis	95
7.1.1	<i>Raw survey data</i>	96
7.1.2	<i>Calibration</i>	99
7.1.3	<i>Optimal coordinates for factor analysis</i>	101
7.2	Factor analysis	103
7.2.1	<i>Model confirmation</i>	103
7.2.2	<i>Factor loadings and descriptions</i>	103
7.2.3	<i>Factor evaluation</i>	108
7.2.4	<i>Article-level and demographic analysis</i>	108
7.2.5	<i>Summary</i>	110
8	Factor narration and conceptualisation	111
8.1	Narrated factors	112
8.1.1	<i>Factor 1 - Journal prestige</i>	112
8.1.2	<i>Factor 2 - Clinical guidelines</i>	114
8.1.3	<i>Factor 3 - Referencing behaviour</i>	115
8.1.4	<i>Factor 4 - Method section</i>	117
8.1.5	<i>Factor 5 - Subjective quality</i>	118
8.1.6	<i>Factor 6 - Basic to applied</i>	120
8.1.7	<i>Factor 7 - Author</i>	121
8.1.8	<i>Factor 8 - Citation meaning</i>	122
8.1.9	<i>Factor 9 - Citation quality</i>	123
8.1.10	<i>Factor 10 - Innovation stunt</i>	123
8.1.11	<i>Factor 11 - Skepticism</i>	124
8.1.12	<i>Factor 12 - Propriety</i>	125
8.1.13	<i>Summary</i>	126
8.1.14	<i>Assessment of participants' articles</i>	127
8.2	Conceptual model	131
8.2.1	<i>Citation impact</i>	136
8.2.2	<i>Journal Impact Factor</i>	137
8.2.3	<i>Citation paths</i>	138
8.2.4	<i>Health and policy effects</i>	140

8.3	Summary.....	141
9	Discussion.....	143
9.1	Research question summaries.....	143
9.2	Implications of results.....	148
10	Final summary.....	151
	References.....	153
	Appendices.....	167
A	Interview data.....	167
A.1	Pilot study transcriptions.....	167
A.1.1	Interview reference: P01.....	167
A.1.2	Interview reference: P02.....	173
A.1.3	Interview reference: P03.....	179
A.1.4	Interview reference: P04.....	179
A.1.5	Interview reference: P05.....	185
A.1.6	Interview reference: P06.....	189
	References, pilot study.....	199
A.2	Main study transcriptions.....	200
A.2.1	Interview reference: M01.....	200
A.2.2	Interview reference: M03.....	206
A.2.3	Interview reference: M04.....	217
A.2.4	Interview reference: M05.....	226
A.2.5	Interview reference: M08.....	238
A.2.6	Interview reference: M09.....	253
A.2.7	Interview reference: M10.....	262
A.2.8	Interview reference: M13.....	272
A.2.9	Interview reference: M15.....	283
A.2.10	Interview reference: M16.....	294
A.2.11	Interview reference: M18.....	300
A.2.12	Interview reference: M19.....	307
A.2.13	Interview reference: M21.....	317
A.2.14	Interview reference: M25.....	327
	References, main study.....	338
B	Interview invitation.....	341
B.1	Original letter of invitation.....	341
B.2	Translated letter of invitation.....	345
C	Interview documents.....	347
C.1	Interview script, pilot study.....	347
C.2	Interview script, main study.....	348
C.3	Declaration of consent.....	349

D	Initial statement codes	351
E	Initial survey questions	359
E.1	Personal details.....	359
E.2	Statements on research quality.....	359
E.3	Importance of article-specific factors.....	363
F	Final survey	367
F.1	Visual design.....	367
F.2	Drop-down response options.....	367
G	E-mail invitation to survey	381
H	Data cleaning	383

List of Figures

4.1	Model of interactions in medical research. Adapted from Lewison (2004).	56
4.2	A model of scientific communication as a global distributed information system. Adapted from Björk (2007).	57
4.3	Records (all publication types) in the PubMed MEDLINE database published between 1945 and 2010, as of late 2011.	60
6.1	Model of study design	69
6.2	Frequency of response codes ordered by rank of occurrence	78
6.3	Number of unique new codes added by each participant, ranked decreasingly	79
6.4	Correlation between professional experience, measured as years since completion of formal medical education, and age in years. Linear regression is illustrated as a red line.	90
6.5	Number of participants from different countries. Only countries with more than two participants are displayed individually.	93
6.6	Correlation between number of participants per country and the corresponding number of e-mails sent to domains from those countries.	93
7.1	Boxplots showing the distribution of main variables measured on 5-point scales (top) and 11-point scales (bottom)	97
7.2	Histogram of variance ranges for 5-point (left) and 11-point (right) scale variables.	97
7.3	Empirical cumulative density distribution of 5-point (left) and 11-point (right) scale variables. All variables are combined into the two groups, creating a collective overview.	99
7.4	Pure frequencies of response selections for article-specific variables.	100
7.5	Histograms showing frequencies of score ranges for index-calibration (left) and standard-calibration (right).	101

7.6	Empirical cumulative density distributions for all variables combined, using index-calibration (left) and standard-calibration (right).....	101
7.7	Distribution of optimal coordinates (oc) with 250 tests, using 500 repetitions in each test, for both calibration types. To the left, boxplots show the overall distribution with quantile ranges, and to the right, the empirical cumulative density distributions for index-calibration (orange) and standard-calibration (blue) show the exact differences between the resulting optimal coordinates for the two calibrations.....	102
7.8	<i>p</i> -value for testing the number of factors, as a function of the number of factors, using Bartlett's test and varimax rotation. The red line marks the 0.05 boundary.	105
7.9	Dendrogram showing the similarity of observations, as likeness of Bartlett scores. Similarity is shown as distance, i.e. values close to 0 are very similar. Clusters are shown as red outlines.....	109
7.10	Heatmap showing the average Bartlett scores for each cluster on each factor.....	109
8.1	Histograms of index scores for variables associated with factor 1 - Journal prestige.....	112
8.2	Histograms of index scores for variables associated with factor 2 - Clinical guidelines.....	115
8.3	Histograms of index scores for variables associated with factor 3 - Referencing behaviour.....	116
8.4	Histograms of index scores for variables associated with factor 4 - Method section.....	117
8.5	Histograms of index scores for variables associated with factor 5 - Subjective quality.....	118
8.6	Histograms of index scores for variables associated with factor 6 - Basic to applied.....	120
8.7	Histograms of index scores for variables associated with factor 7 - Author.....	121
8.8	Histograms of index scores for variables associated with factor 8 - Citation meaning.....	122
8.9	Histograms of index scores for variables associated with factor 9 - Citation quality.....	123
8.10	Histogram of index scores for variable associated with factor 10 - Innovation stunt.....	124
8.11	Histograms of index scores for variables associated with factor 11 - Skepticism.....	124
8.12	Histograms of index scores for variables associated with factor 12 - Propriety.....	125
8.13	Relationship between citations to selected articles and comparable articles. TC = times cited for selected articles, MCJ = Mean citations to articles of the same type, published the same year in the same journal.	130
8.14	Conceptual model of research quality in medical research, from an evaluative bibliometrics perspective.	132

LIST OF FIGURES

9.1	clinrel variable as function of primary job categories "research" or "clinical practice".	144
9.2	Model of impact types and quantitated elements.	147

List of Tables

3.1	Classification scheme for measurement instruments from (Geisler, 2000)...	38
6.1	Basic statistics on item ratings for the Sternberg & Gordeeva (1996) impact questionnaire items. Means are calculated from ratings between 1 (low) and 6 (high).	72
6.2	List of variables, question wording, response types and data types in the online survey.....	81
6.3	Distribution of gender	89
6.4	Distribution of age and experience among participants. Experience is measured as years since completion of formal medical education. Both variables were recorded as exact years but are here cumulated in five-year intervals.	90
6.5	Primary job category	91
6.6	Primary work place	91
6.7	Distribution of medical specialties among participants. Specialties represented by less than five participants were combined in the 'other' category	92
7.1	Descriptive statistics for main variables. SD = standard deviation, var = variance.	98
7.2	Correlations between factors. Factor numbers are used as labels, as the correlation matrix is symmetrical, only the upper half is displayed.....	104
7.3	Proportion of total variance (Pvar) explained by individual factors.....	105
7.4	Factor loadings for main variables. Only loadings $\geq .20$ or ≤ -0.20 are displayed for ease of reading. Salient loadings are marked with bold text. .	106
7.5	Low- and high-scoring factors in each cluster, using thresholds of ± 0.50 . ..	110
8.1	List of papers collected from main study participants, containing quality-related statements.	127

8.2	List of papers collected from main study participants. PY = Publication year, TC = Times cited in Web of Science, total from publication until retrieval (09 Sep, 2011), MCJ = Mean citations for papers of type=article, published in the same journal, the same year with the same citation window as TC. References in appendix. One paper omitted as it could not be retrieved (M18_1).	129
D.1	Initial statement codes and unique ID's as well as frequency (<i>f</i>) of occurrence	351

1. Introduction

Research quality is an elusive but widely used phrase, used by researchers, policy makers and the society as a whole. But what is research quality? In medical research it could be claimed that there is a connection between the quality of research and the overall effect of said research on the state of public health. If medical researchers produce high-quality research, this should be reflected in better treatment, or policies which result in less illness - but this is a very generalised, cursory view on medicine, health, research and research quality. For instance, much basic biomedical research is not directly translatable to health improvements, but through further research iterations it may result in a new drug or therapy that is clinically tested and shown to cure a specific disease. That could be considered one of the purposes of basic research, but this would also banalise central aspects of the basic research idea, such as pure knowledge generation, while the quality also becomes less apparent. The same is true of many other levels and types of research. The complexity increases even more if we wish to *measure* quality as it is of an intangible, complex nature, containing numerous dimensions. This does not make quality measurement irrelevant, however; funding agencies, government bodies and research managers quite naturally want to invest in the best possible research projects. This is where research assessment plays an important role, as evaluating previous research of a given researcher, research group or university might hint at their potential future achievements. These assessments can be divided into two main classes; peer assessment and metrics. Peer assessment is regarded as the golden standard by many but also critiqued for different types of bias and subjectivity, while metrics are usually regarded as more objective but also unable to grasp the entire picture (e.g. Butler & McAllister, 2009; Clerides, Pashardes, & Polycarpou, 2011; Goldstein, 2011; Kenna & Berche, 2011; Taylor, 2011). One great problem of metrics is that they measure some very specific objects, such as publication or citation counts, which may be related to research quality, and used as proxies thereof, but in no way reflect all dimensions or properties of the entire concept (J. R. Cole & Cole, 1971; Zuckerman, 1987). While this is the case, metrics are a relevant tool for research assessment as it may

quantify certain aspects of productivity or impact in the research society that are hard to grasp and easily influenced in peer assessments. But if we want to achieve a good quantitation of the research quality concept it is necessary to first discover the dimensions and properties, or qualities, of the concept. That is the primary aim of this dissertation. In the following introductory chapter, we will look further at the motivation for this subject and illustrate briefly how we will seek to reach the goal.

The theoretical framework of this dissertation is to be found in the bibliometric field, the science of measuring attributes of documents and very often scientific documents. It is a fairly young scientific discipline, closely related to information science and scientometrics, the science of measuring science. Scientometrics and bibliometrics may at times be regarded as synonymous, and at other times the one may be regarded as a specialised case of the other. However, the aim of each discipline is not always the same. Scientometrics are chiefly concerned with the assessment of science (and technology) using various indicators, such as the funding a university receives, the number of researchers employed at a department or the success rate of research grant applications, as well as historical or predictive studies of science, scientific communication and growth and science sociology. Bibliometrics are concerned with similar indicators, such as the productivity and impact of research departments, but always based on information derived from their published items, such as journal articles and monographs. Bibliometrics may also be used for other purposes however; for instance domain analysis (Hjørland, 2002), thesaurus construction (Schneider, 2004), indexing (Salton, 1963; Garfield, 1964; Schneider, 2006) and information retrieval (White, 2007a, 2007b).

This dissertation is concerned with research performance assessment, i.e. the assessment of how well individual researchers and research groups perform on a global and local scale. As already stated, there are several indicators of such performance, but the focus of this dissertation will be on those related to scientific publishing performance. Put simply, bibliometric methods for research performance assessment are about measuring the publication output of a research unit, and assessing the impact of these in one way or another. As discussed above, the connection between impact and research quality is not necessarily straightforward and impact does not reflect all aspects of research quality. Creating a connection between the measurements and the conceptualisation of research quality is a question of what can be measured, what makes sense to be measured and what the meaning thereof is. The famous quote, attributed to Albert Einstein, "*Not everything that can be counted counts, and not everything that counts can be counted.*", comes to mind in this context and is naturally important to keep in mind when performing bibliometric research assessments. The ability to measure publication counts, citations and other quantifiable aspects of science and research does not necessarily mean it is viable to do so in any given setting.

Nonetheless, when it is interesting at all to use quantitative tools to measure scientific output it is because we can see patterns in the outputs and deduce meaning from these patterns. If all scientists were to produce the same amount of research papers per year and citations were given arbitrarily, a measurement of these units would be

of little consequence to anything, as their interpretation would be just as arbitrary as their nature. The case is the opposite, however; the productivity of researchers and the way citations are received resembles natural laws (Price, 1963), although far from all bibliometric research agrees on the validity of such purely statistical measurements of science (c.f. Gilbert, 1977; Cozzens, 1981; Luukkonen, 1997). These first, broad analyses of science and in particular scientific publications, e.g. by Price, show that science and research may be measured and assessed using metrics and statistics and gave legitimacy to the coming bibliometric field. Since then, the fields of sciento- and bibliometrics have developed a number of metrics, indicators and indices, measuring and assessing various aspects of science and research. This has accelerated, in particular in recent years, where many governments and university directors have implemented local and national assessment exercises and funding schemes based on these. This recent focus on bibliometric assessments of scientific productivity has intensified the debate about what these methods and metrics actually measure, and if it is fair to fund research based on such indicators (e.g. Williams, 1998; Elton, 2000; Clarke, 2005; Kostoff & Geisler, 2007). This is not a new debate for bibliometricians; for some time there has been a discussion about the lack of methodological consensus (Glänzel & Schoepflin, 1994), lack of a common language (Glänzel, 1996), database inconsistencies (Glänzel, 1996; Bar-Ilan, 2008), abuse of metrics (Seglen, 1992; Moed & van Leeuwen, 1995; Seglen, 1997), method redundancy (e.g. Bollen, van de Sompel, Hagberg, & Chute, 2009; Leydesdorff, 2009), validity issues (van Leeuwen, 2008), method robustness (e.g. Lehmann, Jackson, & Lautrup, 2006, 2008) and the various problems related hereto. In addition to the above, there is a debate, or even different paradigms, with regard to what a citation, the main measure of scientific impact, is and how it can be used.

To see this in a broader perspective, we can also ask what research quality is. Often the aforementioned citations to publications are used as an indicator of the quality of the research contained within the publication. This shows that research quality can be operationalised, and in such cases can be well represented by bibliometric indicators (Smart, 2005; Andras, 2011), however; while this may be pragmatic and useful, there are a number of issues with the use of citations as a quality indicator (some of which are shown above), and it should also be clear that citations can only reflect a specific part of the quality and impact of research (e.g. Waltman, van Eck, & Wouters, 2013; Wouters, in press). Providing a better understanding of the different dimensions of the research quality concept should improve our communication about the different aspects of research quality assessments, and thereby also clarify which aspects can be measured bibliometrically.

1.1. Objectives of the dissertation

The main objective of this dissertation is to investigate the connection between the research quality concept and bibliometric methodology. This will result in a better and more explicit understanding of what is being measured when bibliometricians speak

of impact and research quality. In order to achieve this goal, it is necessary to first examine the research quality concept in detail with the goal of creating a conceptual model of research quality, describing the dimensions and their relationships. While the present study will focus on dimensions related to bibliometrics, it is presumed that not all dimensions will or can be related to bibliometrics. It should be stressed that it is not the goal of this dissertation to create a universalistic model of research quality, resulting in a single, unified measure that can describe quality in one simple number. Rather the intention is to describe a range of aspects, which may help us to categorize and describe research according to certain criteria.

While one might attempt to create the abovementioned model for scientific research in general, one should also assume it would result in a very superficial model. This would possibly provide us with a good overview of how research quality is perceived, however; it would likely be a very poor tool for clarifying bibliometric methods applicability for research quality assessments. Instead it appears more prudent to investigate the quality concept for a specific scientific field, and thus be able to concretise the findings of this dissertation for said field.

For various reasons the field of medical research has been chosen as the objective of this dissertation. Medical research is one of the most active scientific fields today, with several hundred thousand publications every year. A model of medical research quality aspects will be useful for a large number of institutions, such as hospitals and universities, but also for organisations such as national health ministries or institutes. Medical researchers are also used to being evaluated, and are quite aware of some of the most common bibliometric indicators, such as the journal impact factor and the *h*-index. Therefore, an investigation into researchers' perception of research quality in medicine should provide a large response basis as well as existing knowledge or conceptions of the subject. In addition, explicit definitions of certain types of research quality already exist in medicine, and research has been made into some of these as well, for instance evidence-based medicine (Glasziou, Vandenbroucke, & Chalmers, 2004), or the direct impact of research on health improvement (Lewison, 2002a; Pons et al., 2010).

1.2. Research questions

In order to address the issues mentioned in the above section on the objectives of this dissertation, three central research questions have been formulated. In the following these questions are presented and described.

Research question 1: How is research quality perceived in medicine? The understanding of research quality is central to this dissertation, and any connection between bibliometrics and research quality naturally depends on how research quality is interpreted and which quality aspects can be identified. Therefore the first research question focuses on gathering data to elucidate this topic. Researchers from within the medical field will be included in the investigation, to provide an empirical foundation for answering the research question. To ensure responses from researchers are as open as possible, the first part of this study will focus on interviewing researchers and gathering input from them in the shape of statements on research quality. Following this, a survey will be conducted to quantify the trust in abovementioned statements among a wider audience of medical researchers. These quantified statements will serve as “aspects of research quality” in the following.

Research question 2: How do the identified aspects of research quality relate to each other? Based on the collected aspects of research quality, a factor analysis will be performed to discover which aspects are related. If composite factors are discovered, the statements will be collected and compared to the original data from the interviews to create dimensions of research quality. By comparing the quantitative and qualitative data, it is our hope that causal mechanisms can be described.

Research question 3: How do bibliometric methods apply to the quantifiable aspects of research quality? It is expected that very different types of research quality will be identified in the course of the investigation, of which only some will be directly related to bibliometrics. Those outside the scope of bibliometrics will only be included in a limited extent, as the aim is to create a connection between precisely bibliometrics and research quality. The final question is concerned with exactly this task, building on the findings from the previous analyses. The process can be divided into two tasks; quantitation and implementation. Where quantitation focuses on discussing how the identified qualities can be quantitated in a meaningful way, implementation is about comparing these quantities to bibliometric units and methods. Units should be considered more important in this particular context, as they precede methods in the path from concept to measurement. Metrics might be included in the discussion as well, as a final tier in said measurement path (this is discussed further in section 3.1), but the primary objective is to create the connection between research quality and bibliometric units. Concrete methods and metrics may then be procured from this.

1.3. Structure of the dissertation

To provide the reader with sufficient background on the topics research quality, medical research and evaluative bibliometrics, the following chapters (2-5) will provide just this, along with a theoretical standpoint for the dissertation. Each of these chapters will contain a presentation of the current research on the topics as well as a discussion of the relation to the research questions.

These initial chapters are followed by a description of the overall study design (section 6.1) and details of the methods applied (chapter 6). As we will be using a three-phase study design, some preliminary results will be reported in this chapter as well, while the main results are presented in chapters 7 and 8. The first of these chapters reports results relating to the first two research questions, while the latter discusses the results with regard to previous research presented in chapters 2-4, thereby providing answers to the final research question. The final chapters of the dissertation are dedicated to a discussion of the implications of the results (chapter 9) and a final summary (chapter 10).

2. Research quality

Over the course of the next few chapters, the different topics of this dissertation will be presented in the shape of both a literature review and discussions of the theoretical background. The chapters are divided into three main topics, relating to the research questions, namely *research quality* as a concept, *measuring research quality* and *research quality in medical research*. In this initial chapter about research quality, we will look into existing definitions and epistemological views on the research quality concept, as well as review the existing research on the topic. Only very few studies have been concerned with precisely the concept and meaning of research quality, however; peer assessment, a fundamental way of judging quality, has been studied intensively and hence a specific section is dedicated to this.

2.1. The research quality concept

The focal point of the entire thesis can be boiled down to research quality as a concept. The final aim, as follows from the research questions, is to investigate connections between this concept and evaluative bibliometric methodology. But in order to do so, a definition of the concept must be given first. “A definition” should here be read as “one” and not “the” definition of research quality. Such a definition must necessarily depend on a definition of what research, knowledge and truth are, and consequently of the epistemological views of an individual, field, school or domain. Implicitly in this statement also lies a belief that this concept is socially dependent, however; while the concept can be considered to be a social construct, this dissertation also operates on the assumption that social constructs are definable, quantifiable, observable objects. The reasoning behind this assumption is further explained in the next chapter on quantitation and measurement. In the current context, the important issue is, that while research quality is considered definable; epistemological and domain-specific differences are expected, both regarding research community definitions and the assessment perspective. To clar-

ify this, we will briefly discuss epistemological views of knowledge and the implications for research quality on the most general level. We will not go into detail with specific philosophies of science, but merely focus on the four main epistemological views on the acquisition of knowledge; either through experience i.e. sensory perception (*empiricism*), a priori or intuitively (*idealism*), as a combination of the sensed, intuitive and the abstract (*rationalism*) or through human construction (*constructivism*). These views on knowledge lead to very different views on research; while the empiricist believes experimental observations to be the *truth*, idealists would state that the truth is rather the researchers mental interpretation of the observations (as any observation is a subjective idea of the world). It should be obvious that researchers of either epistemological conviction would emphasise different aspects of research as qualities; most clearly reflected as objectivity versus subjectivity. Also constructivists differ clearly from these views on knowledge - although some may acknowledge a real world, as the empiricists do, true knowledge about this world can only exist as a human construction and the quality of research becomes yet another thing. In nursing, a field closely related to medical research, but with a greater emphasis on qualitative research, the debate about the differing interpretations of quality in quantitative and qualitative research has been going on for some years (c.f. Rolfe, 2006), and in this dissertation research quality will thus be treated as an epistemologically dependent concept, and consequently the research domain becomes a parameter to be accounted for. The philosophical stance represented in this dissertation will be further elaborated in section 3.1.

While several studies deal with different aspects of research quality, impact, esteem, excellence or other related terms, or at least assumed aspects of these concepts, very few studies look closer at the actual concept of research quality, its interpretation and cognition. The studies most closely related to the present topic focus on *impact* (Sternberg & Gordeeva, 1996), *high-performing researchers* and *research performance as a concept* (Bazeley, 2010). Sternberg and Gordeeva (1996) extracted a number of factors influencing what makes an article have impact in the field of psychology, as perceived by the researchers in the field. These factors were derived from statements by the researchers and evaluated by a broader range of researchers as well. Impact and quality are in no way similar, but they are commonly regarded as related in that impact is seen as a dimension of quality. One may therefore anticipate that factors influencing impact will also have some influence on research quality, and these factors will naturally play a central role in the further investigation of the quality concept (see section 6.1). The study by Bazeley (2010) used a mixed methods design to retrieve descriptions of high-performing researchers as well as different dimensions of research performance. In the present context the latter is more interesting, especially as Bazeley constructed a theoretical model of research performance (Bazeley, 2010, p. 897), as a hierarchical model where performance was seen as the combination of *research activity* and *performing* (dissemination and collegial engagement). Research activity was again seen as a concept consisting of four mutually dependant dimensions; *engagement*, *task orientation*, *research practice* and the *intellectual process*. Each of these dimensions are necessary elements in research ac-

tivity, while only one of the dimensions of dissemination and collegial engagement are necessary for performing. Finally, both performing and research activity are necessary, non-causal conditions of research performance (Bazeley, 2010). Bazeley's model has universalistic elements, but being based on a sample of 295 researchers only from Australia, and without field-limitations, the model is not directly translatable to any field or research culture, nor does it account for epistemological differences. The most important distinction between Bazeley's model and the purpose of this dissertation, however, is the connection to bibliometrics, or the lack of same. Nonetheless it is an inspirational framework to which the results of this dissertation can be set in perspective.

With regard to evaluative bibliometrics, research quality is most intuitively related to citation analysis. Over the years, some researchers have called for a theoretical foundation for citation analysis (e.g. MacRoberts & MacRoberts, 1987, 1989; Cozzens, 1989; Luukkonen, 1997), or a qualitative framework (Rip, 1997), as citation analysis was perceived as a field where metrics were being used and analyses performed without the use of theory - which, according to Goertz (2006) is not uncommon in the social sciences. van Raan (1998b) refuted these claims (see section 3.3 for further discussion of this) based on empirical findings. Van Raan's most important argument is that citation analysis differs from reference analysis - a distinction also used and elaborated by Wouters (1999) and which will also be used here - and that citation analysis can not be based on a theory of why and how researchers cite. Whether it is possible or not to create a theory of citations will not be discussed further here, nor will a potential theory of research quality, as this surpasses the scope of this dissertation. It is however feasible and crucial to conceptualise research quality to create models of interpretation. Gerring (2012, p. 112) even considers concept formation to "*[lie] at the heart of all social science endeavors*". A good conceptual model might provide causal arguments, but just as importantly; A conceptual model of research quality for the use in bibliometrics should serve as an interface between the research community, their cognition of research quality, and well-established, empirically founded methods of research assessment, including but not limited to citation analysis. The construction of such a model will thus be a central aim of this dissertation.

2.2. Research quality and impact assessment

In the previous section the concept of research quality was discussed, and an initial connection to citation analysis was established. The theoretical aspects of this connection are clarified in section 3.3, as we shall first explore the practical applications and aspects of research assessment as well as qualitative assessments. This will be carried out as a literature review of different forms of research assessment studies. This should provide an overview of how research assessment has been conducted until now and illustrate some of the challenges of this discipline, in particular in relation to research quality. The review will include and discern between studies of research quality and impact, and while impact is often used synonymously with citations, it is to be understood

in a more general sense here and throughout the rest of this dissertation. Assessments of impact might be citation analyses, but societal impact or changes in policy making would also fit into this category. The difference between the two classes of assessments reviewed here thus relate to attributes of research (quality) and the effect research has on its surroundings (impact). The review will be divided into a general part, followed by specific sections on different types of peer assessment and the impact concept. In the first part we will look into specific problems in publishing which are related to the research quality concept, such as negative findings, fraud, misconduct and retractions, as these issues are closely related to what could be considered dimensions of research quality.

2.2.1. Negative findings, fraud and retractions

Not all research results in a strong conclusion with results supporting the proposed hypothesis. Some hypotheses turn out to be false, or the results may not be strong enough to conclusively support it. Such cases of negative findings or null-findings are potentially important to the research community; From an ethical viewpoint, studies involving human or animal subjects should not be repeated more than strictly necessary to provide the necessary data to reject or accept the null-hypothesis, and if negative findings remain unpublished there is an increased risk of repeating research on living beings that may be potentially harm- or stressful to the subjects. Additionally, contemporary health research is often resource-heavy in terms of funding and personnel. Additionally, clinical practice relies to some extent on results provided by meta-analyses (see section 4.2) which again rely on the sum of positive and negative findings (Easterbrook, Berlin, Gopalan, & Matthews, 1991). It has been argued that negative results may be the result of type II errors (Freiman, Chalmers, Smith Jr, & Kuebler, 1978) or less than adequate hypotheses (Angell, 1989). Easterbrook et al. (1991) however found evidence that the quality of the study design had no association to the likelihood of publication, while there was a strong bias towards the publication of positive results from a selection of UK-based health research projects. Their findings also suggest that studies funded by pharmaceutical industry may be somewhat less likely to be published than other studies. Fanelli (2011) documented a continuing trend of selection bias towards positive results in most research areas, including clinical medicine. In another study, Fanelli (2013) found that positive studies on average received more citations than negative studies. This trend was especially high for clinical medicine, although highly cited negative findings were also observed. Different reasons have been suggested for the bias, of which perhaps the most concerning are neglected submission of negative results from the authors side and refusal of publishing negative results from the editors side. It has not been possible to find conclusive evidence for which of these factors play the larger role, and it is quite possible that both factors are equally present. Other factors influencing the selection of positive over negative findings may be designs that yield false positives (e.g. Ernst & Lee, 2008), or relying too heavily on *p*-values to decide on the significance of results (Ioannidis, 2005), in both cases leading to a misclassification of

negative findings as positive ones. While it is problematic that negative findings are thus masked, it is positive that they are after all published.

Related to the above is the issue of fraudulent research or scientific misconduct - a topic which has received a lot of attention in the recent years with cases such as that of Jan Hendrik Schön¹. A number of papers have addressed the issue by surveying among authors whether they or their colleagues committed deliberate fraud or other *questionable research practices*. Fanelli (2009) gathered 21 of these studies in a meta-analysis, grouping different categories of fraud and misconduct into the two main categories of fabricating, falsifying or modifying data ("fraud") and questionable research practices ("misconduct", e.g. deliberately omitting references to rivaling research groups). Fanelli found lower occurrences of deliberate fraud (2% for self-reported surveys, 14% for surveys asking about colleagues) than for misconduct (34% self-reported, 72% colleague-reported). The imbalance between the surveys indicates that the self-reported results may not have been truthful, and the true figure is likely to be higher. Fanelli also found that researchers in the medical and pharmacological fields were more frequent in reporting either type of misconduct than other researchers. Steen (2011) approached the problem from a bibliometric angle, looking at the 788 English language papers retracted from PubMed in the period 2000 to 2010. Retracted papers are not necessarily fraudulent, but might "just" be erroneous (not deliberate). For the fraudulent papers, Steen found they were commonly published in journals with higher Journal Impact Factor than the erroneous papers, they were often written by "repeat offenders", and the retraction process was often delayed. Steen also found that researchers from the United States of America published a larger amount of fraudulent papers compared to erroneous papers than researchers in the rest of the world.

In conclusion, a number of very different (mis-)practices among authors can affect the resulting research papers in ways which may give a wrong impression of their quality. Citations and the language surrounding negative results might lead us to believe they are low-quality research while in fact they can be of just as high quality as other research. On the other hand, fraudulent papers may appear as high-quality research on all accounts - including citations - until the fraud is discovered. Such discovery might affect the researchers standing in the community and the paper to become retracted, but citations will likely keep existing and potentially even grow in the future.

2.2.2. Peer assessment

As stated previously, peer assessment is one of two commonly used approaches for assessing research performance. Performance can encompass productivity as well as impact and quality, in various interpretations. It is included in this review to reflect the less metric-based side of research assessment, with the various perspectives on research quality and impact, but also to illustrate the intersections between subjective "qualitative" assessments and metrics. The term peer assessment is used to reflect the different prac-

¹A case in which a prolific and promising scientist was discovered as deliberately fabricating data.

tices and interpretations of activities such as refereeing, peer review and peer ranking. While refereeing and peer ranking are mostly used exclusively with specific meanings, peer review is often used interchangeably in the literature, with both of the other and also separate meanings. Wouters (1997, 1999) used peer review to describe each of the different types of peer assessment in the research cycle. The description of Wouters is very useful for understanding the workings, position and importance of peer assessment in research and when a different term is used here, it is merely to stress the importance of discerning between the different types when discussing their use in research assessment. To ensure rigor, a definition of the usage in this dissertation will be given here. Refereeing is used for the act of reviewing a paper submitted to a journal, usually prior to publication, with the aim of rejecting less important, bad and fraudulent research as well as improving the published research and in some cases ensuring a topical, stylistic or linguistic style. While refereeing in no way is particular for health research, it is a well-established institution in this domain, by many regarded as a necessary component in the scientific system, but also criticised (e.g. Reich, Green, Brock, & Tetlock, 2007; Charlton, 2010). This type of peer review mainly serves as a quality insurance tool, rather than quality evaluation, except perhaps in more recent post-publication variations (e.g. Birukou et al., 2011). There are domain-specific differences in refereeing, in the way it is conducted as well as implications and traditions of refereeing which will also be discussed below.

Peer ranking refers to the type of peer reviewing used in some national research assessment exercises, such as the RAE² or the new REF³ in the United Kingdom, where peers review and grade the excellence, esteem or quality of e.g. a university department. Peer ranking allows assessment of intangible dimensions of research which are typically difficult to measure by other means, but also more vulnerable to subjectivity (c.f. Clerides et al., 2011; Goldstein, 2011; Kenna & Berche, 2011).

Peer review is here regarded as a more general term, used for any type of peer assessment that is not refereeing or peer ranking. Typical examples of such types of peer review are evaluations of grant applications, applications and applicants for tenure, residence or other academic professional positions. Grant reviewing will be included briefly, as there is some research on the connection between bibliometrics and the impact of research supported by or submitted to grant committees.

Peer ranking

One of the best known and studied instances of peer ranking are the RAE and REF in the United Kingdom. The RAEs were performed in 1989, 1992, 1996, 2001 and 2008 (Taylor, 2011) while the REF is planned for 2014⁴. One of the key elements of the RAEs were expert reviews of publications selected for assessment by the universities. In the science, engineering and technology areas bibliometric measurements were

²Research Assessment Exercise

³Research Excellence Framework

⁴<http://www.ref.ac.uk/>

included in the most recent RAE. The peer rankings have been subject of a number of papers discussing their usefulness, bias and other problem areas. Taylor (2011) reviewed the results of the peer ranking scores of three management science fields compared to bibliometric scores in three business & management fields. Taylor found correlation between the two types of scores, however; the explanatory variable was department size, which should have no influence on quality. Taylor assessed bibliometric quality as publications from approved journal lists rather than citation rates which only provides a part of the picture. As a result, Taylor found peer ranking to be superior to bibliometric measurement. Clerides et al. (2011) on the other hand compared RAE peer rankings to an array of independent indicators based on publication counts, PhD dissertations, staff and other scientometric variables, in a field closely related to the aforementioned, namely for economics departments. Clerides found there was a bias against new universities, favouring those universities represented in the RAE panel. While only indicative, this illustrates a danger of nepotism or collegial bias when performing peer rankings.

Kenna and Berche (2011) proposed a field normalisation of peer rankings, allowing comparison of rankings across fields. The RAE was used as data material, and the approach of Kenna and Berche appears superior to using mere means of the RAE's research quality, but is also rather specific hereto. The basic model could most likely be transferred to other research quality assessment approaches that depend (directly or not) on the size of the research groups they analyse. Kenna and Berche included a sophisticated discussion of team sizes and their influence on productivity and quality, much in line with the results of Taylor (2011). A final RAE peer ranking study should be mentioned here, namely that of Goldstein (2011) in which RAE assessments were compared to research grant award grades. Goldstein created a scoring scheme to compare different types of award gradings. He found the correlation between award grades and RAE scores was moderate at best, for the five tested fields (economics, sociology, education, psychology, management and business studies). Goldstein emphasized the necessity of a detailed analysis of multiple-person awards rather than naïve analyses which assign the same award grade to each applicant. Goldstein suggested the use of grant award grades as a supplement to peer rankings as they provide different insights into the research process and activities.

In the Italian VQR research assessment exercise, expert panels have a choice of selecting between citations or expert peer reviews to prepare quality evaluations (Abramo & D'Angelo, 2011). Abramo and D'Angelo (2011) argued that while bibliometric measurement is not perfect and should not stand alone, it is superior to peer ranking in terms of "*robustness, validity, functionality, costs and time of execution*". The use of bibliometrics to inform peer ranking, as has been proposed for the future REF, does not, according to Abramo & D'Angelo provide an advantage. In another paper, Abramo, D'Angelo, and Di Costa (2011) compared the Italian peer rankings of excellence to standardised citation scores. They found that the use of citations is superior to peer ranking, as it uses the entire research output, and not just a few, selected papers (excellence).

Another famous research assessment exercise is that of Australia, which is however primarily using bibliometrics. Derrick, Haynes, Chapman, and Hall (2011) therefore had to use a different approach as the abovementioned in order to compare citation metrics for Australian researchers to peer rankings. A voting approach was used to achieve such rankings, and then used to compare peer rankings of Australian researchers in six public health fields to their citation scores, using four different citation metrics. The votes were given based on how “influential” the researchers were perceived to be in their area. Derrick, Haynes, Chapman & Hall found a mild correlation between the two factors, however, all included indices were based on the *h*-index, and the analysis is therefore rather limited with respect to different types of impact. Also Norris and Oppenheim (2011) studied the correlation between the *h*- and *g*-indices for Library and Information Science researchers with a peer ranking (based on both individual and collective scores, gathered through both the RAE and an online questionnaire). Norris and Oppenheim found some correlation between the peer rankings and the *h*- and *g*-indices. A different perspective on the peer ranking discussion was offered by Lewison (2002b), comparing the esteem of journals to citations for a number of biomedical subfields. The esteem indication was based on a peer ranking of journals and thus illustrates a different bibliometric dimensions than the authors. Lewison found that overall there was a correlation between the two variables, but it was higher for the more scientific subfields included.

To sum up the above, many studies have sought to discuss and assess peer rankings of researchers in particular compared to bibliometric indicators such as publication and citation counts. The moderate, yet positive, correlations could indicate that there is some connection between what peers regard as quality, influence or excellence and bibliometric measures. The League of European Research Universities, LERU⁵, also stress that neither peer assessment/peer ranking or bibliometrics should be applied by default, but always considered carefully, as neither covers all aspects of research assessment (LERU, 2012). LERU however considers peer assessment to be the preferred method, in particular as it is the most trusted by researchers, but it can be supplemented by bibliometrics. The report divides assessments into measurements of input, output and outcome, where the latter two are most useful to assess bibliometrically. Outcome in the LERU report is seen as a collective term for the impact on e.g. society, health and production, all dimensions which are valuable to discuss in relation to research quality.

⁵Universiteit van Amsterdam, Universitat de Barcelona, University of Cambridge, Albert-Ludwigs-Universität Freiburg, Université de Genève, Universität Heidelberg, University of Helsinki, Universiteit Leiden, Katholieke Universiteit Leuven, Imperial College London, University College London, Lunds universitet, Università degli Studi di Milano, Ludwig-Maximilians-Universität München, University of Oxford, Université Pierre et Marie Curie, Paris, Université Paris-Sud 11, Université de Strasbourg, Universiteit Utrecht, Universität Zürich

Refereeing

Refereeing refers to the process of peer evaluation of manuscripts submitted to journals prior to publication. Most scientific journals in medicine have some kind of refereeing process for their original research articles (e.g. R. Smith, 1988; Overbeke, 1999) with some variation on the blinding of author and referee names (Overbeke, 1999) and the quality criteria used. While some journals may publish a large number of submitted manuscripts, e.g. the British medical journal only publishes 7% of papers submitted to the journal during a given year⁶, signaling the importance of a strong referee system.

While peer ranking may be regarded as the final cycle of peer assessment (Wouters, 1997, 1999), refereeing is usually one of the first cycles, possibly the first external cycle. Potential prior cycles could be ethics reviews or grant proposal reviews. Nonetheless, Bence (2004) referred to refereeing as first-level peer review and compared it to second level peer review (peer ranking). Bence found first-level refereeing to be necessary and useful while second-order peer assessment in his view was redundant (the papers had already been refereed) and problematic, citing a number of criticisms (such as bias and cost). There are numerous differences in refereeing across disciplines, for instance in mathematics where some refereeing is done post-publication (Zuccala, 2010). In some disciplines, refereeing is blinded or masked while in other disciplines the names of authors and/or referees is fully disclosed. Also the options for responding to reviews vary from pure accept/reject systems to revisionary systems, the latter being the case in medical research. While there is some variation from journal to journal in the health sciences, there are certain steps and practices which are more common than other. In the following is a description of a somewhat standardised approach with some of the major variations included. The description should serve as an insight into the refereeing traditions of the medical field - an important aspect of the research quality concept.

When submitting a manuscript to a medical journal, the author(s) must enclose a cover letter. These cover letters range from standard forms, confirming that all authors contributed equally to the work (in accordance with ICMJE, 1988) to pre-enquiries about the pertinence and merit of the topic (e.g. PLoS Medicine). The editor(s) of the journal make decisions on the further treatment of the manuscript based on the cover letter or manuscript itself, some journals even use several editorial review steps before further action is taken, e.g. the British Medical Journal where all manuscripts are first assessed by one editor, then assessed by a second editor and finally read by senior editors before they are sent to external referees. At each of these stages the manuscript may be rejected in editorial review in which case the manuscript is very unlikely to be published in that journal. Upon initial acceptance a number of external peers are invited to referee the manuscript. In some cases authors are invited to select referees or suggest referees they would rather avoid, in other cases referees are selected entirely by editors. Most journals use either two or three external referees which are most commonly masked so

⁶see the BMJ webpage under “Is the BMJ the right journal for my research article?”, at <http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/bmj-right-journal-my-research-article>

that the authors will not be able to see who reviewed their manuscript. Whether the authors name is masked to the referees varies more, but it is common practice to mask it. In some cases, the full details are disclosed post-publication. This is the case for a number of the BioMed Central journals in the BMC series, which provide a “pre-publication” section for published articles in which the original manuscript and any revisions to it are available along with referee comments and author replies hereto. Referees write a number of comments to the author(s) and editor, including suggestions to improvements, pointing out errors, missing references, flaws etc., and recommend acceptance, rejection or resubmission. Many journals operate with a range of recommendations of “accept”, “minor revisions”, “major revisions” and “reject”, indicating to the editor how likely acceptance should be and to the author how much work needs to be done. The final decision is however that of the editor. It is possible for papers to go through several revisions before they are accepted and published (and thus considered of certain quality worthy of publication).

Albeit not completely recent, a very comprehensive literature review of studies on refereeing (Overbeke, 1999) has gathered many of the different benefits and shortcomings of refereeing, specifically in the medical field. Overbeke identified a number of biases, of which some may be countered through blinding of e.g. author or referee names. He concluded, however:

“Although we do not know much about the peer review [referee] process, it is the only system we currently have to assess the quality of scientific work before publication.” (Overbeke, 1999, p. 38)

While a number of post-publication refereeing approaches have been developed since (Suls & Martin, 2009; Birukou et al., 2011), it is possibly still true that no other pre-publication quality assessment procedures exist. R. Smith (1988) agreed on a number of the issues summarized by Overbeke, and listed the main issues of refereeing as randomness; institutional bias; bias against the innovatory; lack of definition of a peer; fraud; cost and delay. He suggested other ways of measuring performance, especially bibliometric indicators, patents and esteem measures. He also suggested foresight indicators, such as library reprint orders, research proposals and conference proceedings. In the following the literature on refereeing will be reviewed with a focus on the relationship to bibliometrics. Also some of the criticism and alternatives to refereeing will be reviewed, in order to illustrate both potential and limitations of refereeing as a quality assessment tool.

Two recent studies by Allen, Jones, Dolby, Lynn, and Walport (2009) and Lokker, McKibbin, McKinlay, Wilczynski, and Haynes (2008) compared expert ratings of papers, post-publication, to different impact indicators. While the studies used post-publication refereeing, the setups were attempting to emulate the normal refereeing process somewhat. Allen et al. (2009) thus used expert reviewers to rate the importance

of 687 medical papers supported by the Wellcome Trust⁷. The reviews were performed, at most, six months after publication, divided into categories of importance and compared to citations and F1000⁸ ratings. Most papers were distributed around the category, which indicated a minor advancement in science. Only 20 papers were considered landmarks, i.e. contributions of very major importance. A correlation between Journal Impact Factor and importance rating was found (median Journal Impact Factor increased from 2.31 to 30.12 for the lowest and second-highest importance-rating respectively) and also total citation counts (3-year window) had a positive correlation with the importance rating (median citations increased from 5 to 41.5 for same importances, the highest importance-rating is here omitted due to very few observations, but showed even higher scores). The same was found for F1000 ratings, and the number of authors and the number of citations were positively correlated. Some correlations may be explained by field differences - e.g. author numbers, which may also affect total citations due to self-citations. While Allen et al. found a generally high correlation between qualitative expert reviews and quantitative indicators, there was also large variation on the level of the individual papers. Lokker *et al.*'s study also found a positive correlation between citations (2-year window) and expert ratings of the clinical relevance of medical papers. The authors claimed that a number of paper-attributes, some of which are physical while others are based on expert ratings, allow the prediction of citation rates two years after publication, just three weeks after the paper is published. A more detailed analysis of F1000 ratings and citation correlations was performed by Waltman and Costas (2013), comparing the complete F1000 set of recommendations to citation scores, showing a clear correlation between the scores and the citations given to the papers, even when accounting for field differences. Waltman and Costas found that several highly cited articles were not included in the F1000 ratings, and while the ratings could be used as a predictor of highly-citedness, journal-citation indicators were found to be better predictors hereof.

Using data from the journal *Angewandte Chemie International Edition*, a chemistry journal with high Journal Impact Factor, a series of studies investigated different correlations of actual referee ratings (Bornmann & Daniel, 2008a, 2010b, 2010a). Two of these papers (Bornmann & Daniel, 2008a, 2010b) found that papers submitted to the journal, and accepted by the referees (Bornmann & Daniel, 2008a) or the staff editors (Bornmann & Daniel, 2010b) on average receive more citations than papers which were rejected but published in other chemistry journals. In another paper (Bornmann & Daniel, 2010a), refereeing was used as a predictor, and the validity thereof was examined. The conclusions were not as clear here, although refereeing was generally found to be a predictor of paper quality. Some of the uncertainty may be due to interaction between initial rejection reviews and subsequent improvements to those papers when submitted to other journals.

⁷The Wellcome Trust is a UK-based, global charity, funding health research. See: <http://www.wellcome.ac.uk/>

⁸F1000, Faculty of 1000, is a post-publication refereeing organisation, focused on identifying excellent research in biology and medicine. <http://f1000.com/>

Although a paper by Zuccala (2010) focused on the refereeing practices in mathematics, which are quite different from those in medicine, there are certain suggestions here which might be transferable to medical research and potentially provide new assessment tools in line with some of the suggestions of Lewison (2002a, 2003)⁹. A post-publication reviewing system exists in mathematics where referees are asked to write short reviews of papers that have been published, helping readers to decide whether they will read the paper or not (Zuccala, 2010). A somewhat similar system can be found in a few medical journals, such as the *Annals of internal medicine*, where editors provide the reader with a (short) description of the relevance of the study. While the main concern of Zuccala (2010) was another, the study suggests connections between citations and post-publication assessments. This could be compared to editorials, invited letters to the editor and letters in general in medical research. Potentially these document types may be used to discuss other aspects of impact than traditional citations, an issue which is also taken up by van Leeuwen, Costas, Calero-Medina, and Visser (2012) though not specifically for medical research.

When talking to researchers, one will often hear dichotomous statements about refereeing, such as “it is vital” and “it is flawed, referees never agree”. The latter is clearly illustrated, albeit in a fairly old example, by Starbuck (2006) who in 1968 as editor of *Administrative Science Quarterly* found almost no correlation at all between referee scores: “It was so low that knowing what one reviewer had said about a manuscript should tell me almost nothing about what a second reviewer had said or would say” (Starbuck, 2006, p. 19). This type of subjectivity in refereeing expresses something about the applicability (or lack of same) of refereeing for research quality assessments. While somewhat anecdotal, Starbuck’s findings are reflected in more recent studies on the limitations of refereeing; some are even more critical, pointing out crucial flaws. One of these flaws is the hypothesis that papers on an important topic are more likely to be accepted for publication. This hypothesis was somewhat, although not completely, debunked by Reich et al. (2007). They looked at the leniency of referees, in terms of the likelihood of a referee accepting flaws in a manuscript and allowing publication despite these flaws. Their hypothesis was that a) if a paper is on an important topic, referees will also consider the paper to be more important, or b) they may, unintentionally, be less careful. Also Type I and II errors were considered with respect to importance. They found that in particular the latter affects the leniency of referees while a) and b) hardly have any effect. A number of other flaws have been mentioned above, such as randomness, different types of bias, fraud, cost and delay (R. Smith, 1988; Kassirer & Champion, 1994; Overbeke, 1999; Rennie, 1999; Suls & Martin, 2009). Most recently, the various bias types have been analytically reviewed (Lee, Sugimoto, Zhang, & Cronin, 2013) with the aim of discussing impartiality in refereeing. Lee et al. found that most studies indicate a bias, but limitations in the studies along with arguments of the social nature of refereeing lead to a more fundamental question of whether impartiality should actually be an ideal for refereeing or if not certain types of bias are actually

⁹see chapter 4 on page 48 for more details on this

a positive effect of refereeing, as it leads to improved efforts from the author (Lee et al., 2013). On the opposite end of the scale, a very harsh critique of the refereeing system can be found from Charlton (2010). Charlton went so far as to compare refereeing to a “parasitic cancer” of bureaucracy which has infected the scientific system and, in his views, must be removed. Charlton also observed other problems in the scientific system and referred to it as “zombie science” - something which is dead on the inside but keeps moving anyway. While some of Charltons statements were likely meant as provocations against the system, editors, funders and research managers, he echoed some of the same critiques as above.

Subsequently it is not surprising nor uncommon to see Winston Churchills famous quote on democracy rephrased as “peer review is the worst system for scientific publishing - except for all the others [own wording]” (e.g. Nature neuroscience editors, 2005). The difference being that not many other systems for scientific publishing have ever existed, nor have they been thoroughly tried. Some alternatives exist though (e.g. Suls & Martin, 2009; Birukou et al., 2011), but while they exist, only small steps have been made at adapting them. Some approaches stand out as having some success, such as the open review model adapted by some BioMed Central and Public Library of Science (PLOS) journals. Other approaches also exist, such as the arXiv in physics, Open Exchange and the Adversary Model. In the latter the referee becomes the authors’ adversary in a more equal position, similar to the roles of prosecutors and defendants in law (Suls & Martin, 2009). Also community-based approaches exist, primarily focusing on post-publication assessment, featuring e.g. bidding, rating or voting systems. While there is some opposition against these systems, fearing that bad research may be promoted by just a few people, Birukou et al. (2011) argued that bad research is more likely to be ignored, given the little time researchers have, for performing such tasks anyway. If bidding is implemented pre-publication, there may be other implications with regard to the speed of bidding, and the potential of judging papers to be more important if the title, and possibly abstract, is written in a manner to promote the importance of the topic, regardless of the remaining content of the paper (Birukou et al., 2011). This falls somewhat in line with Reich et al. (2007) who found that the risk of referees being too lenient with papers on important topics was greater, if the assessed text was short. All in all, refereeing is viewed as one of the cornerstones of the scientific system without which a lot more irrelevant, faulty or fraudulent research would be published, despite the shortcomings, biases, randomness and flaws of the system itself. While intrinsically related to research quality it should be kept in mind that refereeing is a quality *insurance* tool and not an *assessment* tool.

Grant review

Grant reviewing differentiates itself from peer ranking and refereeing in that it is both a quality insurance and assessment process, and potentially a quality indicator as well. As grant reviewing takes place prior to manuscript submission, there is no direct connection between citations and grant reviews, as can also be seen in Wouters (1997). There is

however the possibility that successful grant applications result in better research. This hypothesis was approached by Lewison and Dawson (1998), with the formulation that funding from several sources indicates especially important research questions. Lewison and Dawson compared the Journal Impact Factor of journals in which different studies with funding acknowledgments were published. They found a positive correlation between the number of funding bodies and the average journal impact factor and while this methodology has some limitations (e.g. it is unclear if the papers were among the highly cited papers in the journals), it indicates that more funding leads to more impact. This is seen as a causal relationship, as the more funding sources also means more reviews, which should improve the quality of the research conducted.

Another study showed that certain bibliometric indicators might predict the outcome of fellowship applications (Bornmann, Mutz, & Daniel, 2008). While the study's focus was an analysis of variants of the *h*-index, their methodology allows the above conclusion to be drawn. Bornmann, Mutz and Daniel compared grant reviews conducted prior to the invention of the *h*-index to different *h*-index derivations for the applicants. They found that especially the *m*-index correlated very well with the acceptance and rejection rates of a particular review board. In a different study investigating the bias of a grant committee towards gender, nationality and previous success rates Bornmann and Daniel (2007) found no bias and a positive correlation between citations and the acceptance and rejection rates for grant applications, somewhat strengthening the findings of Lewison and Dawson (1998).

Grant reviewing does not play as major a role in bibliometrics as publication-based studies, but it is essential to much medical research in order to secure funding for often expensive trials and experiments. Subsequently there may be an even larger potential in this area for biblio- and scientometrics, despite skepticism as expressed by Goldstein (2011).

2.3. Summary

The traditional approach to assessing research quality has been, and still remains, peer assessment of some type. However, as has been illustrated in this chapter, all types of peer assessment, be it ranking, refereeing or grant reviews have problems or limitations which increasingly lead to an inclusion of e.g. bibliometric evaluation as a supplement to the more subjective peer assessments. The relevance of documenting the relationship between research quality and bibliometrics thus becomes increasingly important as the use of bibliometrics increases. This is especially true as bibliometric studies are not always being conducted by experts, but are more often being executed by “amateur” assessors (Gläser & Laudel, 2007).

We have also seen how several factors “behind the scenes” of the published document, such as fraud, play a role for the quality of research, and how negative findings can be very important, high-quality studies, despite not leading to e.g. new treatments in medical research. Despite this fact, several negative results remain unpublished and

less cited than their positive counterparts. While neither peer nor bibliometric assessment can lead directly to the publication of negative results, we can hope that specialised studies of these might help emphasise their importance and role in scholarly communication. This further illustrates the need for an articulation of what research quality is. In the next chapter we will follow up on this by looking at concept construction and quantitation, with regard to research quality and bibliometrics.

3. Measuring research quality

The title of this chapter might appear provocative to some; how can we postulate to measure something as intangible as research quality? And how can bibliometrics contribute here? Typically bibliometricians will refer to measurements of impact, a concept also related to research quality. The discussion of this relationship is precisely summarised by Gläser and Laudel (2007):

“The reconciliation between the complex and indeterminate concept of research quality and the measurable property of impact has been achieved by regarding impact as *an aspect of* research quality that can be measured by bibliometric indicators, and that bibliometric indicators can therefore serve as ‘partial indicators’ of quality” (Gläser & Laudel, 2007, p. 103)

However, there is no further definition of what this aspect of research quality is, what other aspects there are, and how impact can be seen in relation to other aspects of research quality. Part of the explanation for this problem is that research quality to some extent is domain-dependant (as also argued in the beginning of the previous chapter), but as the present dissertation operates in a specified domain, this problem is thus minimised. In addition the dissertation is written from a realist viewpoint, more specifically scientific realist, meaning that the premise for the following chapter is that we believe a real world exists and that scientific experimentation can provide us with the best possible approximation of the truth about this real world, while keeping in mind that any theory about the world is an approximation hereof, that it is fallible, and thus may be improved over time (Bunge, 1993). This conception of a “real world” should not be confused with a denial of the existence of social constructions. Research quality, as a concept, is a theoretical construct (as is impact), and measuring a construct as complex and intangible as research quality does not appear feasible or valid, however, this does not mean that the concept can not be quantitated. There is a fine but very important distinction between measuring and quantitating, and in the following chapter arguments for the quantitation of research quality are posed. This is followed by a

view on measurement theory and specific implications for evaluative bibliometrics, and specifically citation theory and analysis, and their relation to research quality.

During the course of this chapter some central terms will be used synonymously in order to signify different meanings of the same basic entity, and to allow distinction between different entities using the same term in the literature. In order to minimise confusion, a short account of this is given here. The terms *concept*, *construct*, *entity* and *object* are all used about research quality, as a concept to signify its nature as an abstract object; as a construct to signify that it is an idea (a real and observable one) but not a physical object, as an entity to illustrate its independent existence, and finally as an object in the scientific understanding, of something that is observable, and the object of our study. *Attributes*, *dimensions* and *qualities* are used interchangeably about the properties of concepts, constructs etc., as the term *quality* is the most precise in the context of quantitation, but might be confused with the use in the phrase *research quality*. These terms, as well as *properties* should be seen as completely synonymous in the current context.

3.1. Quantitation & measurement

As briefly noted above, according to Bunge, quantitation and measurement are not the same, although some may refer to them as the same, and while the difference may appear trivial, important differences and causal relationships are implied. When this distinction is essential to this dissertation, it is because of the underlying nature of the concepts, as described below. Quantitation precedes measurement, and conceptualisation precedes quantitation. If we wish to measure research quality, we must first create the concept and theoretically sound quantitations of the concepts qualities. The following review of literature on quantitation and measurement will elaborate on this viewpoint.

The interpretation of the abovementioned concepts is here based on Bunge (1999) who describes measurement as a "*laboratory or field operation*" while quantitation is purely conceptual (Bunge, 1999, p. 84). Quantitation is thus a theoretical or even philosophical act and it precedes measurement. Whether it is possible to measure or quantify objects is highly dependant on the object in question but also the property we want to empirically investigate. Following Bunge, if we observe the temperature of a cup of tea, it is a measurement, while if we ask a number of people to indicate on some scale how well they like tea, this is no measurement. By best possible design it will be ordinal values which will provide us with some insight, and a quantity. Measurements can thus be mathematically superior to quantities, as they allow us to perform a wider range of mathematical and statistical operations than possible on the ordinal (or nominal) scales. This also allows us to induce laws from measurements. This contrasts strongly with relativist and constructivist views on concepts or constructs, which would rather emphasise the ideas behind and communication about the concept. While these are valid approaches as well, and might indeed offer additional insights into some di-

mensions of research quality, measurement is not their primary concern. As our aim is to link research quality to the measurement-heavy discipline of bibliometrics, it appears prudent to maintain this scientific realist view.

3.1.1. Quantitation of quality

For anything to be measured, it is an essential requirement that it is quantifiable. In the social sciences, different approaches have been made towards a philosophical theory ranging from positivist to postmodernist or social constructionist (e.g. Lundberg, 1939; Kemper, 1981; Bryman, 1984; Bunge, 1999) with fundamental consequences for the perception of social phenomena (Kemper, 1981) and sociological methodology (Bryman, 1984). Bunge (1999) proposes a connection between sociology and philosophy which builds on quantitation and allows for measurement. Bunge's philosophy is strongly opposed to the relativist approaches, which Bunge describes as anti-quantitative. Much like e.g. Sokal (Weber, 2007), Bunge's philosophy has a realist worldview, and provides a theoretical basis for empirical quantitation in the social sciences. While this may be opposed by postmodernists, Weber (2007) in a defense of realism argues that the opposing relativist "*conceptualization of science as mere social construction is misleading and even dangerous not only for science itself but for the whole society*" (Weber, 2007). As stated above, the dissertation at hand was written from a scientific realist viewpoint, with Bunge (1993, 1999) as the main theoretical and philosophical basis for quantitation and measurement theory. Bunge is chosen, as the scientism aspect of bibliometrics as a social science, fits well with his focus on quantitation in social science research. Quantitation is at the heart of Bunge's philosophy; everything can be quantified, except for existence itself:

"I submit that only one property is, with all certainty, intrinsically qualitative, namely existence. I also submit that in every other case quantitation depends exclusively on our ability and interest, so that in the face of failure to quantitate we should suspend judgment and encourage others to try."
(Bunge, 1999, p. 75)

When Bunge speaks of quality and quantity, it is important to note that Bunge does not perceive these concepts as exclusive but as mutually complementary (Bunge, 1999, p. 71). The way Bunge defines these concepts, and the way in which they will be used in this dissertation, quantity depends on quality, as qualities define the objects we wish to quantify. A set of objects may thus be defined by the qualities they share; if for instance we wish to measure the citation impact of female, Norwegian physicists¹⁰, we look at articles authored by researchers sharing certain qualities; they are female, from Norway (defined either by birthplace or more likely institutional affiliation), publishing in the Physics field and possibly other qualities as well, such as a certain minimum number

¹⁰A hypothetical study almost identical to an existing study of Aksnes, Rorstad, Piro and Sivertsen (2011).

of publications within a given timeframe. All these qualities are shared by the objects under investigation, and they can be quantified as either the number of objects or the intensity of one of their common qualities. These quantities may be of a purely quantitative nature (e.g. age or salary), in which case we can measure them (see below, section 3.1.2). If however we are dealing with ordinal values, such as self-assessed research experience (e.g. *none, some, extensive*), where one may consider a magnitude and a natural order, and thus apply numbers to values (1 = none, 2 = some, 3 = extensive), these are not true quantities, but semi-quantities (Bunge, 1999). Numbers used to describe semi-quantities are not real numbers but rather numerals or ordinal values. It is important to understand that numerals cannot mathematically be treated the same way as real numbers and therefore semi-quantities cannot be considered real measurements either. If we consider the above example with the cup of tea, we can express the temperature very precisely and e.g. plot the cooling over time, or measure the average temperature of tea water when brewed in a specific machine. However, if we collect opinions such as 0 = “I don’t like tea”, 1 = “I like tea”, 2 = “I like tea very much”, an average value of these numerals will not tell us very much. If half of the respondents dislike tea and the other half like it very much, an average value would lead us to believe everybody likes tea somewhat. These simple observations can be conferred to more complex constructs as well, as any construct can be described by its qualities. This interpretation of a construct, or concept, as an entity possessing a number of qualities, might appear as an essentialist interpretation, in which essential qualities uniquely define the concept. This is however not the intention - while conceptualisation is also regarded as description through qualities, the existence of other definitions and qualities is acknowledged. The most important element here is to identify the qualities as they are (latently) perceived by the research community.

As has been mentioned before, and based on Bunge’s view on objects, qualities and quantitation, it follows that quality precedes quantity - there can be no quantity without prior consideration of quality, albeit in some cases trivial.

With regard to the main topic of this dissertation - *research quality* - the goal is to identify qualities, or attributes, of the concept which can be said to determine “something” about research quality. In line with the preceding argumentation, no a priori knowledge about the concept is assumed and inductions should be empirically based. The empirical input for the conceptualisation of research comes from previous research, as described in chapters 2, 3 and 4, as well as a combined qualitative and quantitative study of medical researchers’ perception of research quality, detailed in chapter 6 and reported in chapters 7 and 8.

While this dissertation should be seen in the light of scientific realism, there are epistemological and ontological views of the world, in which the concept of research quality would be something entirely different. From a relativist perspective, there is no real world, but only our idea of the world we experience, and quantitating research quality is equivalent to objectifying something they would consider to be purely subjective. Also pure positivists might argue against the concept of research quality as it can

not be experienced, and social constructivists might argue that it would not be possible to clearly define a dynamic, socially dependant concept before it has changed. Antirealist positions have been taken by proponents of each of these epistemological views, referring to e.g. the failure of Einsteins theory of quantum mechanics versus that of Bohr¹¹, as an implied failure for realism. Bunge (1993) as well as McMullin (1984) make strong cases for why this is not a defeat for scientific realism; McMullin by showing how Bohr's theory of quantum mechanics is no less realist than Einstein's - albeit further removed from the deterministic classical mechanics, and Bunge by pointing out that failure is a step towards better representations of knowledge: "... scientific error is corrigible in principle: that we can frequently go from error to partial truth to better approximation" (Bunge, 1993, p. 231). This view is reflected in this dissertation, and consequently our results are not expected to provide a universally, globally, infinitely true definition of the research quality concept, but the best, current approximation based on the available data, parameters and assumptions. From a scientific realist viewpoint, this is the best we can hope to achieve, and a necessary step in the progress of science. With this statement we conclude the theoretical background and justification for the quantitation of research quality. The following section elaborates on the next logical step: measurement, specifically with regard to bibliometrics.

3.1.2. Measurement & metrics

Metrics, such as bibliometrics, may generally be defined as:

"a system of measurement that includes: (1) the item or object that is being measured; (2) units to be measured, also referred to as 'standard units', and (3) value of a unit as compared to other units of reference." (Geisler, 2000, p. 34)

From this follows that quantitation conceptually precedes measurement in that an object or quality cannot be measured unless it has been quantitated. The additional requirements for metrics imply that not all objects and qualities are necessarily measurable, and more importantly, while quantitation may be fairly straightforward for a concept such as *scientific impact*, the operationalisation as a measurement may be considerably more complex. A simple, illustratory - but thematically unrelated - example: if we wish to find out, how many passengers fit on a given airplane type, we can easily measure this, simply by counting seats inside the passenger cabin. This measurement system is not necessarily the only relevant one though, as we could instead allow passengers to stand up, and measure their weight in order to fit as many passengers as possible. Other, more scientifically related, examples can be found e.g. in physics, where different systems of standard units require some conversion, e.g. from degrees Fahrenheit to Celsius or Kelvin, or from kilowatt-hours to joules. When it comes to the social sciences - and bibliometrics - neither the objects, standard units or reference units are

¹¹the "Copenhagen interpretation"

necessarily as straightforward or precise as those used in the natural sciences. Firstly, research objects in the social sciences may be diffuse and the qualities we wish to measure may be abstract or otherwise hard to quantitate, for instance if as peer reviewers we are asked to grade on a scale from 1 to 5 how relevant the research presented in a paper is. Secondly, the objects in question may actually be processes rather than states or magnitudes, in which case “[w]hat we don’t know about such phenomena - and sometimes what we find so difficult to measure - is precisely what we wish to measure” (Geisler, 2000, p. 35). Geisler (2000) and Bunge (1999) thus agree on a very central subject; in order to measure something, first we need to know what we want to measure - then we can find out how.

Bibliometrics is no exception from the above. The field differs from a number of other social sciences in that the predominant measurement instrument (citation and publication counts) is adherent (see table 3.1) and therefore in theory relatively unobtrusive, relying on bibliographic databases, and therefore maintaining a degree of objectivity. The unobtrusiveness is only a partial truth; while the instrument does not directly influence the observed measurements at the time of observation, bibliometrics may have an impact on the behaviour of researchers, research policy or funding (see sections 2.2 and 3.3). While this has no effect on the objectivity of the instrument, it does have statistical consequences and raises ethical questions (e.g. Moed & van Leeuwen, 1995; Cronin, 2000; Yu, Yang, & Liang, 2009; Weingart, 2011; van Raan, 2006; Falagas & Alexiou, 2008). This is however not the subject at this point.

	unobtrusive	intrusive
Passive	a. ADHERENT Advantages: (1) relative objectivity (2) ease of use	c. OBSERVER Advantages: (1) variety of viewpoints (2) knowledge about the phenomenon
	b. FOLLOWER Advantages: (1) versatility of design (2) ability to compare	d. CONFEDERATE Advantages: (1) direct measurement of on-going items (2) ability to measure processes

Table 3.1: Classification scheme for measurement instruments from (Geisler, 2000)

The type of measurement instruments utilised in bibliometrics almost exclusively rely on bibliographic databases; be it citation or publication counts, co-citation or co-author analyses, bibliographic coupling or other types of relationships between words, documents, authors, subjects and institutions. Generally they can be divided into two main groups with respect to measurement: counts and relations. The objectivity of the counts has a dichotomous nature - at the time of recording¹² the bibliometric researcher does not influence the object of study through their observation, however, the reported results might subsequently affect the behaviour of researchers, politicians, funders or

¹²e.g. data collection from a citation database

research managers. The recording in itself is thus objective and unobtrusive, however; the recorded data might have been affected by previous analyses and thus carries with it a degree of subjectivity. Given the accuracy of bibliographic databases, however; these can be very accurate instruments. Geisler (2000, p. 42) lists a row of traits which are all considered highly effective for bibliometric counts: accuracy, precision, repeatability, tolerance, bias and sensitivity. Geisler does not relate to bibliometric networks, or relations between bibliometric units. These can be considered a type of process and as such on a higher level of abstraction than mere counts.

In the above, research quality and other composite concepts have been referred to as being complex or abstract, whether in terms of quantitation or measurement. In the context of this dissertation, research quality is to be regarded as a construct of several factors or aspects. Each of these factors may, in a measurement terminology, be regarded as dimensions, and research quality may thus be regarded as a multidimensional concept. According to Geisler (2000), the path from such *conceptual constructs* to an actual measurement is long: we must first select appropriate dimensions, then variables and indicators and finally measures. This model illustrates the layers of bibliometrics well, and also serves to place the subject of this dissertation in a context. Symbolically it operates with the interactions between conceptual constructs \Leftrightarrow dimensions \Leftrightarrow variables, by focusing on the dimensions and relating these to the concept and the available variables. We can create a logical order of the abovementioned terms into a general model, illustrating the understanding of concepts, quantities and so forth as they are used in this dissertation:

concept \rightarrow *qualities* \rightarrow *quantitation* \rightarrow *variables* \rightarrow *units* \rightarrow *metrics*

, the first half of this model illustrates *conceptualisation* and the latter half *measurement*. The first two research questions of this dissertation aim at a conceptualisation of research quality, including a quantitation hereof, while the third research question concerns measurement, albeit on a theoretical level. The measurement part of the dissertation is related to evaluative bibliometrics as well as indicator theories. Both topics are treated below.

3.2. Evaluative bibliometrics

In the above we have seen that units and variables are essential elements in any type of measurement, and in the following sections we will therefore focus on bibliometric units and variables, as the basis of bibliometric measurement. The initial section will give a very brief introduction to the foundation of bibliometrics, which is followed by an account of specific bibliometric units and measurement issues related to these. While some of these might appear trivial to the experienced bibliometrician, the theoretical understanding of the different units is essential to the coupling of concept and measurement. Citations, as a specific unit, are singled out and treated individually, as this unit has been specifically linked to scientific impact, as an aspect of - or proxy for - research quality (J. R. Cole & Cole, 1972; Zuckerman, 1987; Gläser & Laudel, 2007).

This is diverted into two parts of the current section (parts 3.2.3 and 3.2.4) concerning the implications of citation analysis and the impact concept, as well as the following section on indicator theories (section 3.3).

3.2.1. Bibliometric foundation

As was briefly mentioned in the first chapters of this dissertation, bibliometrics has a foundation in statistical analyses of the quantitative attributes of documents, as discovered by e.g. Price (1963). In the following section this will be further elaborated as it plays a critical role for the understanding of the following sections on bibliometric units as well as citation, referencing and indicator theories. The early 20th century introduced the first serious quantitative and statistical approaches to literature analysis, markedly by George Kingsley Zipf and Jean-Baptiste Estoup who were the first to discover statistical regularities in language, the English and French respectively (Estoup, 1916; Zipf, 1949; Harremoës & Topsoe, 2006). Zipf introduced what is referred to as Zipf's Law by counting all unique words in James Joyce's "Ulysses" (Joyce, 1922) – a plot of ranked word frequencies revealed an inverse power law, which has been shown to be approximately true for all natural language, albeit with some limitation due to the finiteness of natural language (Tripp & Feitelson, 2007). Other prominent research on the countable properties of text was conducted by Alfred James Lotka and Samuel C. Bradford, showing the same type of distribution as in Zipf's Law but for the number of papers produced by different authors in a scientific field (Lotka, 1926) and for the distribution of papers published in scientific journals (Bradford, 1985). These very similar distributions are often referred to as Pareto principles or 80-20-distributions (e.g. 80% of papers are written by 20% of the authors in any field). Also citation distributions have been shown to be described by power laws (e.g. Price, 1963; Katz, 1999), and in a manner resembling natural laws.

These discoveries of fundamental properties of language and publishing have laid the foundation for modern information retrieval (IR) as well as bibliometrics – two closely related disciplines. The vector space model created by Salton (1971) and the term clustering methods (Spärck-Jones, 1971) used in IR rely heavily on Zipf's research, as much as the first bibliometric research relied on Lotka's and Bradford's findings (Price, 1963).

While this early bibliometric research was of a very quantitative nature, not all researchers in the area of research assessment have agreed on this approach, applying more sociological or linguistic approaches to the assessment of science. The epistemological consequences of these different approaches are vast, and for that reason the differing approaches to research assessment will be presented, by discussing what has been phrased as "theories of citation" (e.g. Leydesdorff, 1987, 1998; Moed, 2005), thereby illustrating the known connections between citations and research quality. As this term has been used to cover both statistical and sociological approaches, despite certain crucial differences in the interpretation of *citations*, the sharp distinction of Wouters (1999) will be used, in which *indicator theories* are used to cover different interpretations of

referencing, citations and semiotics. In the current context the use of “citations” relates to statistical implications of citation aggregates, while the discussion of a theoretical understanding of citations is reserved for later.

3.2.2. Bibliometric units

Units are traditionally understood as a magnitude of some quantity, and this is also the case in bibliometrics where authors, papers, citations, journals etc. are considered units. However, bibliometric units are also related to each other, which increases the complexity of bibliometric units compared to simple counts or measurements. Additionally, different units may operate on different levels of complexity, for instance; a journal consists of several papers and a paper has one or more authors, while a paper is (generally) only published in one journal but an author may have written several papers. All these relationships are thus not equal, some are one-to-many and others are many-to-many, and operating on these different levels of units leads to different levels of complexity (Egghe, 1998). Before we look further into measurement issues, a few paragraphs will be used to explain this complexity, as it is the basis of unit aggregation - a central element in statistical citation analysis.

To start with the beginning, the primary object of bibliometric studies, and the source of the available units of analysis, is the document. In most cases, documents will be publications i.e. published documents, and bibliometric analyses will be restricted to document representations in shape of metadata records for these publications¹³. From the publication as object follows a series of units allowing measurement of different attributes of the document on different levels. These attributes are typically authors, journals or aggregations thereof, e.g. all journals in a subject category or all authors affiliated to a specific university. For an in-depth description, please refer to Vinkler (1988). These units have varying relationships with the source document, with different complexities. E.g. an author may - at the time of publication - belong to just one institution or several, but a country will always contain multiple institutions and authors. Articles typically appear in only one journal, and if they should appear in several journals, they are likely to be counted as several articles, maintaining a 1:m relationship. While complexity can increase with aggregation, it increases even more when citations and references are used as units of measurement, rather than mere productivity (Egghe, 1998). In addition to this attribute of citations, there is a small, but important, difference between citations and referencing (Price, 1970; Egghe & Rousseau, 1990; Wouters, 1999) which will be discussed further in section 3.3. With regard to bibliometric research assessment, primarily the publication and citation studies are of interest. While many of these units appear self-evident and straight-forward, there are theoretical issues concerning most of them. In the following we will briefly discuss these, as they are of consequence to the understanding of how bibliometric measurements may be undertaken.

¹³The term “publication” will be used for both actual documents and document representations in this dissertation, as the distinction is of little consequence in the specific context.

Authors

While authors are ordered arbitrarily or alphabetically in some fields, medicine has a tradition for ordering authors by contribution; the first author tends to have contributed the most, and the last author is often the senior researcher or supervisor (Baerlocher, Newton, Gautam, Tomlinson, & Detsky, 2007). This information may be used to make detailed analyses of researchers and research groups, for instance, the supervisor of a research group might be identified as the most commonly occurring last author of a coauthor analysis, or to identify the “actual” authors of a paper featuring numerous co-authors. In addition to these norms, most medical journals adhere to the Vancouver Declaration (ICMJE, 1988), which defines what an author is, hereunder a requirement of explicit contribution to all phases of a research project. Despite these traditions and regulations, it has been shown that far from all authors live up to the full list of criteria in the Vancouver Declaration (Bates, Anić, Marusić, & Marusić, 2004), and that there is large variation in the strictness of authorship rules within the domain. Many medical journals now demand explicit declarations from authors, either in line with the Vancouver Declaration or as an extension thereof (Hudson Jones, 2003) to reduce the problem. When this is a problem in the first place, it is due to the use of publications for career advancement, and the thereof following incentive for claiming authorship of papers that one has only contributed marginally to (Flanagin et al., 1998; Bennett & Taylor, 2003; Gøtzsche et al., 2007; Wislar, Flanagin, Fontanarosa, & DeAngelis, 2011) or even scientific misconduct, claiming famous authors as co-authors, stealing research or hiring ghost writers (see e.g. Strange, 2008).

Also with regard to authors as a bibliometric unit, the above domain specificities are extremely important factors for how one may assess research output, what should be taken into consideration, how results can be interpreted, and what consequences results have for aggregated levels of information. Attempts to draw conclusions on an institutional level based on the productivity of authors employed at this institution is thus highly dependent on how authors are counted, and how they are aggregated.

Documents

Documents are very diverse entities, not only when it comes to their content, but also their type, source and other bibliographic parameters can be important factors for bibliometric assessment. Concerning the document type, which is likely the most prominent factor for research assessment studies, it may be regarded on two levels; the literal document type and the study type. Document types are typically monographs, anthologies, journal and conference articles, but other types also exist. Journal articles can be further divided into reviews, letters, editorials, research articles etc. Finally, these documents may be divided into different study types, which will be discussed further in section 4.2 on evidence-based medicine. The reason for including this discussion of document type classification here is the consequences it might have for bibliometric assessment. Counting documents can be extremely difficult, as different types require very

different amounts of effort and are used very differently by the scholarly communities. Whereas journal articles typically represent comparable work effort, e.g. monographs require a somewhat longer commitment. Some documents, like books, are likely to become study material and might have a wide use and impact although without receiving many citations. Such phenomena have been observed as differences in citation characteristics for articles, books and book chapters (Botts & Hargens, 1991). But also factors tied to metadata attributes of documents have been found to be related to citation rates of journal articles (MacRoberts & MacRoberts, 1996; Bornmann & Daniel, 2008b) in a degree that might be considered a systematic error in citation studies (Waltman, van Eck, & Wouters, 2013). These factors are in particular number of co-authors and impact of references (Bornmann & Daniel, 2008b).

Words

The most essential content of textual documents is the words present in the publication, and their meaning. These words and phrases, as well as words found in the title, keywords and subject headings can be used bibliometrically (e.g. Schneider, 2004). In information retrieval research and natural language processing, these are the key elements, however, in bibliometric assessment of research they are rarely used. A famous exception to this is the context-based citation analysis of Small (1978), revealing insight into citing behaviour. Co-word analysis has also been used to analyse sociological aspects of research (e.g. Callon, Courtial, Turner, & Bauin, 1983; Leydesdorff, 1989), but with clear limitations (Leydesdorff, 1997). Thus, despite the apparent value of document contents for the actual advancement of science, it is of fairly limited use for bibliometric research assessment.

Journals

Journals can be seen as an aggregate of documents, and while other aggregations exist, e.g. anthologies and conference proceedings, journals play the major role in medical research. It is commonly accepted that some journals are more prestigious than other, and publishing an article in a highly esteemed journal may be more difficult due to stricter policies or a larger selection base. This is often compared to differences in citation rates for journals, which can be seen e.g. in the Thomson Reuters Journal Citation Reports (JCR) or the Scimago journal indicators. These journal indicators are based on citations to papers within the journal, and the aggregates attempt to reflect a utility, interdisciplinarity, impact or other attribute of the journal. For some indicators, this has been shown to be problematic, e.g. the journal impact factor which is equivalent to average citations to papers, while this distribution is highly skewed (Seglen, 1992). A plethora of other issues have been identified for journal indicators (e.g. Moed & van Leeuwen, 1995, 1996), but that is not the focus here. Journals may be further aggregated into subject categories, as is the case in JCR. Also this type of aggregation can be considered problematic, in particular with respect to defining clear categories of journals and the

criteria defining such categories (Pudovkin & Garfield, 2002; Leydesdorff & Opthof, 2010a; Moed, 2010; van Raan, van Leeuwen, Visser, van Eck, & Watlman, 2010).

Units summarised

We have seen that the various bibliometric units are all derived from one central object, the publication, that they are interlinked and highly complex. We have also seen that some of the units are indirectly linked to dimensions of research quality. This link should become more evident when studying citation analysis and especially impact.

3.2.3. Citation analysis

Whereas the other units described above can be directly extracted from document representations, as they are attributes of documents, the nature of the citation is different. Only when a document is seen in relation to other documents the citation becomes apparent. Citations are the aggregated references to one document from other documents, and whereas the references can be seen as unique, static attributes of each document, the citation aggregate is dynamic, as it is susceptible to change whenever a new document is published. This property makes the citation conceptually different from the other objects, and also more complex (Egghe, 1998). We will here review some critical aspects of citation analysis; focusing on the relation to research quality. This review can be divided into demographic factors, self-citations and negative citations as well as measurement issues.

Several studies have shown that certain demographic data may be associated with citations, for example the nationality of authors. This has been shown by e.g. Campbell (1990), Grange (1999), Cronin and Shaw (1999) and West and McIlwaine (2002). Also gender, age and professional position has been investigated by e.g. Aksnes, Rorstad, Piro, and Sivertsen (2011). While this study is based purely on Norwegian researchers, it finds that female researchers are cited slightly less than their male counterparts, taking fields or journals into account. When taking the Journal Impact Factor, of the journals they publish in, into into account, the difference is very small, as female researchers tend to publish in journals with a slightly lower Journal Impact Factor than male researchers. They also find that female researchers publish somewhat less - in particular, female researchers are less likely to publish more than just a few papers. Additionally, a large difference can be observed for career tracks - almost as many female PhD students graduate as male ones, but far less female researchers are employed in the higher tiers of the research system (Aksnes et al., 2011). This could potentially be combined with a generation effect. Also Landes and Posner (2000) investigated the age effect. They found that age is a positive factor for receiving citations, especially cumulatively, but that the positive effect of age depreciates at a certain point.

Other than demographics, an issue that is of great importance in citation analysis is the act of self-citation i.e. an author referencing their own work (author self-citation), or references in a paper to other papers appearing in the same journal (journal self-

citation). While there are issues with the latter, these are mostly political and we will here focus on author self-citations. When they are problematic it is because authors can reference their own work for various reasons - either in a legitimate way in order to e.g. build on previous work, or in order to promote themselves through irrelevant self-referencing, boosting their citations and visibility disproportionately. Self-citations can be excluded from citation analyses, which however might be misleading in the first case. A very extensive study was performed by Bonzi and Snyder (1991), asking authors for their reasons for self-referencing. In most cases, authors reference their own work for legitimate and professional reasons, and for almost all the same reasons as they refer to other researchers, although there is an increased tendency for referring to own work in order to establish one self in a field, as well as to build on previous work. References to other researchers are more common, when the purpose of a reference is to demonstrate knowledge of important work in the field (Bonzi & Snyder, 1991). In a different, large-scale study Aksnes (2003) found that self-citations are received earlier in the process than other citations, and while they grow over time, the growth rate is smaller than that for other citations. The share of self-citations will thus decline over time, and varies according to scientific field. van Raan (1998a) found that studies with international collaboration receive more self-citations than other studies, although the increase is not as large as the overall increase in citations for internationally authored papers. Van Raan therefore argues that self-citations should be accounted for in citation analyses.

Very similar to self-citations are the journal self-citations, i.e. citations from one journal article to another published in the same journal (usually within the previous two years). Just like author self-citations this is a natural consequence of the structure of the scientific communication; when one publishes articles in the journal most relevant to the research contained therein, there is a high degree of probability that same journal will contain other articles of relevance for citing. However, when journals strategically use self-citations it becomes problematic. An example of such strategies, albeit on a small scale, can be seen in Opatrný's paper on the protest of *Folia Phoniatrica et Logopaedica* against the "reign" of the journal impact factor (Schutte & Svec, 2007; Opatrný, 2008). Policies intended to boost citations by demanding authors to cite papers from said journal (R. Smith, 1997) and thus increase the journal impact factor or the immediacy index, are even more problematic. To counter the effects of such policies, some citation-based assessments also exclude journal self-citations (Yu et al., 2009).

Also negative citations are debatable; should they be counted at all, or perhaps even subtracted from citation counts, as their intention is to demote and not promote the research in the publication referred to? Discerning between positive and negative citations is a very labour intensive task and none of the existing citation databases feature this. However, this might not be a problem. First of all, the fraction of negative citations is almost negligible (Garfield, 1979), and from this may further be concluded that;

“If scientists tend to ignore inferior work that is of little importance, then the work that they do go to the trouble of formally criticizing must be of some substance.” (Garfield, 1979).

It is further argued by Garfield that the importance of negative citations depends on the purpose of the citation measurement; whether it is a matter of “being right” or if it is a matter of influence on the scientific community (impact). As is the case with most research assessments and in particular those of interest to this dissertation, citations are used exactly to describe the impact of a given publication; a numerical indication of how many other publications have been influenced by said work. The intention of the citation thus becomes less relevant.

For both citations as well as authors an issue rises on the ground of the aforementioned many-many relationships. If an article has more than one author, should each author receive full credit for that paper, or should it somehow be divided among the authors, as it could be assumed that the work behind had been divided as well? This is why different counting methods have been implemented, which can be applied to either publication counts for authors, or citation counts (or other impact indicators). The earliest examples of these counts used somewhat different terms, such as whole counts, fractional counting or complete counts. As a response to this, a nomenclature was suggested by Gauffriau, Larsen, Maye, Roulin-Periard, and von Ins (2007), in which counting methods are ordered in counting classes. The methods are denoted as complete, straight and whole either with or without normalisation. Their paper is also a good source for a number of mathematical operations which are relevant for the different counting classes and methods. In a second study, some of these counting methods were compared with respect to differences in ranking of countries. They found that countries with high levels of international collaboration benefit from whole counting (all unique countries collaborating receive equal credit) (Gauffriau, Larsen, Maye, Roulin-Periard, & von Ins, 2008). Also Huang, Lin, and Chen (2011) used five different counting methods to rank countries with regard to publication counts as well as citations. For paper counts, rank changes were fairly small, for some citation counts, the rank changed somewhat more. While rank changes appeared small, they were all highly correlated. They conclude that the rank changes were fairly small although some individual countries showed large variations. A different study ranking countries by field-normalised citations, using different counting methods, showed a clear benefit for countries with high degrees of international collaboration when they used whole counting (Aksnes, Schneider, & Gunnarsson, 2012). Their findings thus resembled those of

Gauffriau et al. (2008), with some additional methodological strength, compared to Huang, Lin and Cheng due to the use of field-normalisation.

The reliability, validity and robustness of different citation-based indicators and indeed bibliometric measurement itself has been a subject for some time, but still receives some attention. The Journal Impact Factor has received much attention (e.g. Seglen, 1992; Moed & van Leeuwen, 1995; Seglen, 1997; Greenwood, 2011; Falagas & Alexiou, 2008; Larivière & Gingras, 2009) and more recently also the *h*-index has been widely discussed (e.g. Lehmann et al., 2006; Bornmann et al., 2008; Costas & Bordons, 2008; Lehmann et al., 2008; van Leeuwen, 2008). These studies focus on various issues of the indicators, e.g. the lack of robustness of the *h*-index. A complete review of problems with different indicators would be extremely lengthy and while there are many important points in these studies their primary application is beyond the scope of this review. With regard to measurement however, the study by Ibáñez, Larrañaga, and Bielza (2011) shows that a number of bibliometric indicators are connected through dependencies which can be mapped through Bayesian networks. These networks show what we might expect, based on the aforementioned complexity discussion, namely that the more complex, derived indicators depend on the more simple, pure indicators. Using their model, it is possible to explore these dependencies. Other studies have also investigated the correlations between indicators, e.g. Bollen et al. (2009) who investigated 39 impact measures through Principal Components Analysis. They found a number of different factors, and the principal components explaining most of the indicator differences were time (“rapid” versus “delayed” indicators) and reputation (“prestige” versus “popularity”). Leydesdorff (2009) also compared a number of bibliometric indicators using factor analysis, identifying two main factors, namely “size” and “impact”, explaining a very large part of the variance between indicators. More recent developments attempting to counter some of the measurement problems of bibliometric indicators have focused on the use of percentiles rather than scores or ranks (e.g. Lehmann et al., 2006, 2008; Leydesdorff, Bornmann, Mutz, & Opthof, 2011; Andersen, Christensen, & Schneider, 2012). Another recent development in measurement practice has introduced statistical significance testing (e.g. Opthof & Leydesdorff, 2010), which on the other hand is opposed by Schneider (2013) who argues that the use of significance testing comes with a danger of over- or misinterpreting results, especially when randomness can not be assumed (as is the case in most bibliometric studies).

This section of the review indicates that there are many different measurement issues that need to be considered for bibliometric measurement. A very important issue is that bibliometric units are not as straightforward and more complex than many other physically occurring units, such as those used in classical mechanics (e.g. measuring speed as meters traveled per second). This leaves room for both interpretation, variation but also the danger of misinterpretation.

3.2.4. Impact

The concept of impact is often, especially in bibliometrics, used as synonymous with citations. While impact is sometimes seen as a dimension of research quality, it is important to note that it cannot be substituted, and while quality carries a positive meaning, impact may also be seen negatively as it basically signifies “influence”. Also medical researchers will often know this interpretation, in particular as the journal impact factor. However, impact can be used to signify numerous other meanings, e.g. the effect of science on society (and vice versa), political impact, or more intangible effects on research other than the direct, measurable references. In medical research, impact also has the specific interpretation of health effects, including number of patients potentially affected by a new treatment and the size of the effect. The term impact by itself is therefore highly ambiguous if taken out of context and should always be treated accordingly, which will be attempted in this dissertation by not using the term without accompanying qualifiers, such as citation impact or societal impact, unless referring to the very general, overall concept. In the following review, impact in a number of different interpretations will be discussed, in particular of course with regard to its very close relationship to the quality concept.

Impact is of special interest in research assessment as it differs from a number of other overall variables, such as research activities and output, in that it is not controlled by the evaluated unit (Kostoff & Geisler, 2007). While a researcher may take steps to improve the impact of their research, such as publishing in high-profiled journals or promoting their research at conferences or online, the impact of their research ultimately depends on its reception in the surroundings, e.g. by inclusion in reference lists, or as a change in clinical practice. No matter which of these types of impact one decides to assess, it is still of great importance to consider the effect assessment might have on the assessed unit, even if it is out of their control. This is also stressed by Kostoff and Geisler (2007), insisting that assessors and research managers take the time to select proper assessment methods, corresponding to the intended goals of the assessment, and the strategic goals of the research unit, rather than “cherry-picking” among available methods. If failing to do so, it may have unintended consequences for the involved units (Kostoff & Geisler, 2007).

A very comprehensive entry point for impact studies is offered by Kuruvilla, Mays, Pleasant, and Walt (2006), describing different types of impact observed in health research projects and gathered in the Research Impact Framework. Impact is here divided into research-related impact, policy impact, service impact and societal impact, which are detailed further in their paper. Some of the detailed impact types could also be considered dimensions of a research quality concept, such as impact on health status and evidence-based practice or translatability potential.

Also factors pertaining to bibliometrics are mentioned, such as publications and papers and research networks. Citation impact has been studied from several perspectives, including the connection between authors own perceptions of importance and impact. In two different studies Aksnes (2006) and Aksnes and Rip (2009) look into

the connection between how important researchers consider their own papers to be, and the citations they receive (Aksnes, 2006) as well as the importance of the citations they receive (Aksnes & Rip, 2009). The studies were performed on the same researchers, and in the first study they found papers considered to be major contributions, by their own authors, to receive more citations in general than those considered to be just minor or medium contributions. The findings thus correspond well with the previously mentioned study by Allen et al. (2009), but with the important addition that there was large variation for individual articles, thus indicating that citations correspond well with the importance of studies, if seen across a collection of papers, but not necessarily so well if only a few papers are analysed. In the latter study by Aksnes and Rip (2009), the researchers were asked how important they considered citation counts to be and what their perception of the meaning of citations was. While the researchers agreed on many views, of which many were also in accordance with bibliometric findings, they also found that:

“The shared repertoire may be too cynical, and lacking empirical support. It continues to be reproduced, however, because negative experiences tend to be more salient (and there is little opportunity to get a view of patterns at more aggregate levels), and there are no incentives to check the validity of the repertoire views.”(Aksnes & Rip, 2009)

This quote illustrates the very different perspectives bibliometricians and individual researchers have; while the bibliometrician operates on the aggregate level, the individual researcher observes the citations given to their own or other individual papers. A few aggregates, although in a different sense than in the above, are known to some authors, e.g. the Journal Impact Factor, reflecting citations to a journal. This indicator is quite well known among medical researchers, and the image on the aggregate level also appears to be less cynical than for individual papers, as researchers actually believe the Journal Impact Factor to reflect the quality of journals (Sønderstrup-Andersen & Sønderstrup-Andersen, 2008). The belief in the Journal Impact Factor as a quality indicator is also reflected in a study by Perneger (2010), comparing citations to different publications of four consensus statements. These statements (QUOROM¹⁴, CONSORT¹⁵, STARD¹⁶, STROBE¹⁷) were published in multiple journals, and hence with different journal impact factors. Perneger found a positive correlation between citations to the statements and the journal impact factor of the journals they were published in. This clearly indicates, albeit from a small sample, that authors are more likely to select references from journals with higher journal impact factor.

Suggestions have been made by Lewison (2002a, 2003, 2004) that bibliometrics might benefit from incorporating different document types, such as e.g. newspaper

¹⁴QUOROM: QUality Of Reporting Of Meta-analyses

¹⁵CONSORT: CONSolidated Standards Of Reporting Standards

¹⁶STARD: STAndards for Reporting of Diagnostic accuracy

¹⁷STROBE: STrengthening the Reporting of OBServational studies in Epidemiology

articles, policies, regulations and textbooks to discover more about other types of impact than just citation impact, in particular in medical research. As Lewison notes;

“In some ways bibliometrics is at the stage of European navigation in the middle ages. The familiar territory is well, even obsessively, charted but beyond the known world there are only unknown dragons on the map.”(Lewison, 2002a, p. 191)

Also S. Smith, Ward, and House (2011) note the problem of focusing unilaterally on one impact type, here in particular the upcoming Research Excellence Framework assessment in the UK which only focuses on research impact. Smith et al. also criticise the limited involvement of metrics in the same context. Health impact has been investigated and compared to citation impact by e.g. Pons et al. (2010), who compared mortality rates for congestive heart failure at 50 Spanish public hospitals to citation rates of papers published in cardiovascular disease research at the same hospitals. They conclude that there is a low to moderate correlation between the two variables but that teaching and technology level at the hospitals is of greater importance than the research impact.

The final part of this review will be dedicated to a new array of units coming from the altmetrics community, which has been developed mostly as an alternative to the Journal Impact Factor (Priem, Piwowar, & Hemminger, 2012). The indicators operationalise some of the same units as bibliometrics, e.g. journals and authors, but introduce new units from the Internet and in particular social media sites. Examples of units are downloads, views and readership (Priem et al., 2012; Wouters & Costas, 2012), and while these add new perspectives to scientometric assessment, they also come with new problems of interpretation and standardisation (Wouters & Costas, 2012). Arguably the most important feature of altmetrics is the diversification of “impact” and allowing analyses of unintentional influence on e.g. practitioners or the general public, or the use of informal publishing sources (e.g. blogs and other social media), thereby allowing insight into the more invisible scientific culture which traditional bibliometric units can not analyse. Also Parinov and Neylon (2011) stress the importance of new assessment practices to match the changes in science, e.g. through new interpretations of impact.

From the above it is clear that there are different perceptions of impact in health research. While research and citation impact are well known, health impact plays a major role, and each impact type is viable for consideration in relation to research quality.

3.3. Referencing and citation theory

Citation analysis encompasses a number of very different approaches to quantitative studies of research (e.g. Moed, 2005; De Bellis, 2009) with the one commonality that they study citations given from one research paper to another. While this might seem rather straightforward - and technically it is - we will see below that there is far from consensus on the meaning of citations or the act of citing and what one may deduct

from empirical analyses thereof. We will also see that there are bridging approaches which allow us to operationalise different ontological views on citations, in particular that of Wouters (1999). Firstly, we must however look at the differing viewpoints, often referred to as the normative approach and the social constructivist approach. We will also include a more statistical approach as it was suggested by van Raan (1998b). The purpose of this section is to document these different approaches, as the relationship between citations and research quality must necessarily be affected by the theoretical framework around citations.

Citation analysis plays a central role for research assessment, as it is often used as a proxy for scholarly impact, an expected dimension of research quality. Although a select view citation studies (De Bellis, 2009) were conducted previous to the creation of the Science Citation Index in the 1950's and -60's, most theoretical work on citations started with the work of Eugene Garfield (Garfield, 1955, 1979; De Bellis, 2009), resulting in the construction of the aforementioned SCI. One of the most important intents of the SCI was its use as an information retrieval tool (Garfield, 1964), where the citation is regarded as a subject entry point, much like keywords or descriptors, a view shared by Salton (1963) and somewhat hinted at in the visionary article by Bush (1945). The idea was developed further by Small (1978), regarding references as symbols of a concept. The reference symbolises a meaning deduced by the author from the referenced document, in a given context, marked by the sentences immediately surrounding the reference in the specific use. Small's view of references is perhaps the most comprehensive, but also the most labour-intensive, as the meaning of each context must be deduced separately. Small developed the concept of *uniformity* to somehow compensate for this, allowing a citation analyst to classify contexts into specifically (pre-)defined categories of meanings, thereby also allowing analysis of changes in reference use over time (Small, 1980). More recently, the use of citation contexts has been used to create semi-automatic indexing of cited documents in a specific setting (Schneider, 2006).

The idea of using citations for statistical analysis of science sprang from these early works and was famously developed further by Price (1963, 1965). Garfield and Price were not only concerned with the opportunities of citation indices and the natural-law-like phenomena of scientific research, they also reflected on the social phenomena resulting in these types of behaviour and the underlying structures and mechanisms of scientific communication, however; the first approach towards a theory of referencing in a sociological context came from Robert K. Merton (Merton, 1942, 1968, 1973). Merton's viewpoint was a normative one; the act of citing is a social act, in which researchers adapt to the norms of their field through schooling and communication. The result is that they use references as a "reward", giving them to those who came before them as a recognition of their work. Norms also influence which references are used; if several references could be used in a given context, the researcher will adhere to the norms of their field and use the most commonly used reference, thereby resulting in a

Matthew-effect¹⁸ (Merton, 1968). Merton's views were somewhat supported by those of Kaplan (1965), who used initial findings from the SCI to support the assumption of norms in science and diverging norms across disciplines and countries. Also S. Cole and Cole (1967); J. R. Cole and Cole (1971) further developed the idea of a normative approach to the meaning of references, and were perhaps the first to draw a connection between citations and research quality. Both the Coles and later Zuckerman (1987) regarded citations as a (measurable) proxy for direct expressions of research quality.

In contrast to Merton's normative view of science stands the idea of using references as linguistic tools to persuade the reader (Gilbert, 1977). The use of references becomes an argument and their use does not necessarily depend on any type of norms. References are not used to give credit, or document the research coming before one's own, although there may be an overlap between references used as persuasive arguments and those documenting the heritage of one's ideas. Citation counts in this perspective are measurements of how many authors considered the article useful as a rhetoric tool. Cozzens (1981) argued that references could be given as reward as well as for persuasive matters, but that the rhetoric side would always weigh more heavily. More recently Luukkonen (1997) also defended these views, including the theory of Latour (1987). The key belief in these theories is that persuasion is a stronger force than reward, that research is a social system and that the meaning of references is constructed socially and used as tools of power in a communication process - hence it is also referred to as the social constructivist approach to citation theory.

Of the various consequences of the different approaches, the most polemic is the Ortega hypothesis, first postulated by J. R. Cole and Cole (1972) and subsequently confirmed by other researchers (e.g. White, 2004; Bornmann, de Moya Anegón, & Leydesdorff, 2010) and refuted by others (e.g. MacRoberts & MacRoberts, 1987, 1989). The Coles formulated the hypothesis based on the thoughts of the philosopher José Ortega y Gasset, that the progress of science rests on a mass of mediocre research(ers). The Coles challenged this idea, claiming research progresses by the work of the elite, based on the references given by top-researchers to other top-researchers - as would be the result of the normative approach in its most pure form. The consequence would be, in the most extreme interpretation, that most mediocre researchers could be fired at no harm to the progress of research. This is opposed by MacRoberts and MacRoberts (1987), supporting the views of Gilbert. They supported this empirically in another paper (MacRoberts & MacRoberts, 1989), by analysing how several facts which (in a normative view) should have been cited, remained uncited. White (2004) and more recently Bornmann et al. (2010) somewhat confirmed the views of Merton and the Coles, claiming that reference profiles of top-cited research(ers) were more heavily based on top-cited research than mediocre research. White also added that Gilbert as well as the MacRobertses show good examples of normative behaviour in documenting the

¹⁸An expression adapted from the biblical Gospel of Matthew: "For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken even that which he hath." —Matthew 25:29, King James Version.

research they build on themselves. Both White and Bornmann et al. did not claim reward to be the only explanation, however; but it explains their findings more exactly than persuasion would, and it does not “require a tangle of motives” as White stated persuasion would.

The gap between the two perspectives on the meaning of references also results in a difference in the meaning of citations. While these perspectives are commonly referred to as theories of citation (c.f. Cozzens, 1981; Leydesdorff, 1987) it would be more exact to label them theories of referencing (Wouters, 1999; Moed, 2005). A completely different approach to citation theory, which actually focuses on the citations rather than references is proposed by van Raan (1998b), stating that the various differences in reasons for using references even out when citations are aggregated as is the case in most bibliometric research assessment. Van Raan made a comparison to physics; studying referencing in order to create a theory of citation is equivalent to studying individual molecules to create a framework of thermodynamics (van Raan, 1998b). It could be added that parallel to Van Raan's analogy, there are branches of physics and chemistry that are dealing with the study of individual molecules, revealing other properties than studies of ensembles of molecules (e.g. Moerner & Orrit, 1999; Zhuang et al., 2000). The analogy thus holds for bibliometrics and the most important observation of Van Raan is that the unit of analysis pertains to the level of observation, and so must the theory. Also Egghe (1998) supported this view for bibliometrics, showing that the complexity of different types of citation analysis depends on the level of observation and the types of objects studied, corresponding quite well to the above.

Rather than regarding citation and referencing theories as mutually exclusive or competing, Wouters (1999) grouped the various theories into a *science of science* cluster and a *sociological* cluster of indicator theories. One of the implications was an articulation of the two clusters, and also the very different levels of observation in these. This perspective thus integrates well with van Raan's, while leaving room for other views too. Wouters also included two other theory clusters (a semiotic and an information science cluster) which are not expected to play a role in the current context and thus not treated further here. The science of science cluster is equivalent to a strict interpretation of theories of citations - represented by e.g. van Raan (1998b) - while the sociological cluster corresponds to theories of referencing. This duality of references and citations can be understood in a semiotic framework, where it becomes apparent that there are decisive differences between the units (Wouters, 1999). This view is supported by Cronin (2000) who developed the semiotics analogy further. Creating this distance between citations and references might help create the understanding of their differences and at the same time allow the gap between the approaches to be bridged.

Both citations, as citation aggregates on the item-level as well as the container-level, and references as social, communicative objects, are linked to research quality in some way. Following the work of Merton (1942, 1968, 1973), S. Cole and Cole (1967); J. R. Cole and Cole (1971), Zuckerman (1987), White (2004) and Bornmann et al. (2010) it appears reasonable to see the act of referencing as related to an expression

of quality; by referring to a particular work, the author acknowledges previous work (*reward*) and as a peer assesses the quality of this work. While this may not represent the entire image of referencing, there is strong evidence suggesting that this type of normative behaviour is clearly present (White, 2004; Bornmann et al., 2010). The link between references and research quality is thus related to whether authors consider references to be viable, relevant research. If medical researchers act according to the principles of evidence-based medicine, one would expect this to be true. The link between references and research quality should thus be strongly related to the evidence concept. This evidence-link should arguably be even more apparent for citations and research quality, as citations are aggregates of references. One of the key arguments in the science of science cluster of indicator theories is that the random effects on referencing behaviour caused by individual preferences etc. in the sociological cluster cancel each other out when regarding citation aggregates, much like the random, local behaviour of single gas molecules statistically can be summed into the laws of thermodynamics (van Raan, 1998b). If indeed a link between medical evidence and quality perception in referencing behaviour exists, this link should become even more apparent when aggregated, if it is a strong link.

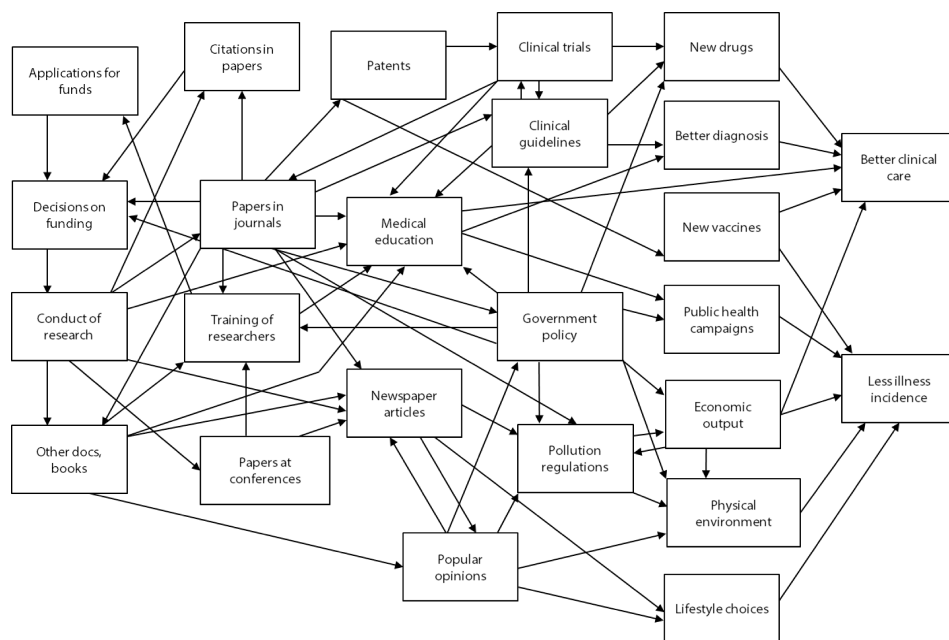
In conclusion, several approaches to indicator theories exist, and while there is a large gap between some viewpoints, the work of Wouters bridges this abyss and allows us to consider references and citations as supplementary sources of information. With regard to research quality, it is here proposed that the concept can be seen as the third angle in the triangle; meaning that neither references, citations or research quality are identical, but each concept is clearly linked to the other.

4. Research quality in medicine

In some areas of medical research, an important element in getting useful results is the definition of proper endpoints; if for instance a new cardiovascular drug is tested, endpoints could be defined as atrial fibrillation, cardiac arrest, absence of recurring symptoms or mortal outcome. The same types of endpoints are naturally important when discussing the quality of medical research and ultimately the goal of all medical research is better treatment and/or less incidence of diseases. A very broad definition of medical research quality could thus be health improvement caused by research. While one could argue how this statement justifies any medical research, as it has an inherent quality, and assessments of research thus are superfluous, the process is not that straightforward in reality. In addition to obvious cases of fraudulent, medical research (Steen, 2011), the research community is experiencing a tremendous increase in the amount of papers, while many papers remain infrequently cited or even uncited (Stern, 1990; van Leeuwen & Moed, 2005), but also an increasing number of papers which are published only to further careers (Reidenberg, 1989; McKibbin, Wilczynski, & Haynes, 2004). But even health research conducted for the “right” reasons, using the “right” methods and delivered in a timely and precise manner may be several steps from actual health improvement. While a clinical study of a new drug may lead to immediate improvements in health for a specific group of patients, a number of studies must be conducted prior to this; on the origins and pathology of the disease, interactions with drugs, proteins or other types of treatment, different stages of the disease and so forth. But also after the conduction of such a clinically relevant, successful study it needs to be adapted into practice, e.g. by inclusion in a clinical practice guideline. While one might potentially observe a direct relationship between some types of clinical research (equivalent to applied science) and health improvement (this is discussed further below), the relationship is far less obvious for some other types, and even less for biomedical research, as this discipline focuses on basic research in health. Also, research is rarely a straightforward process; the path from discovery of a new biomarker in a genetic research lab to the development and clinical testing of a new drug based on those basic research results may be

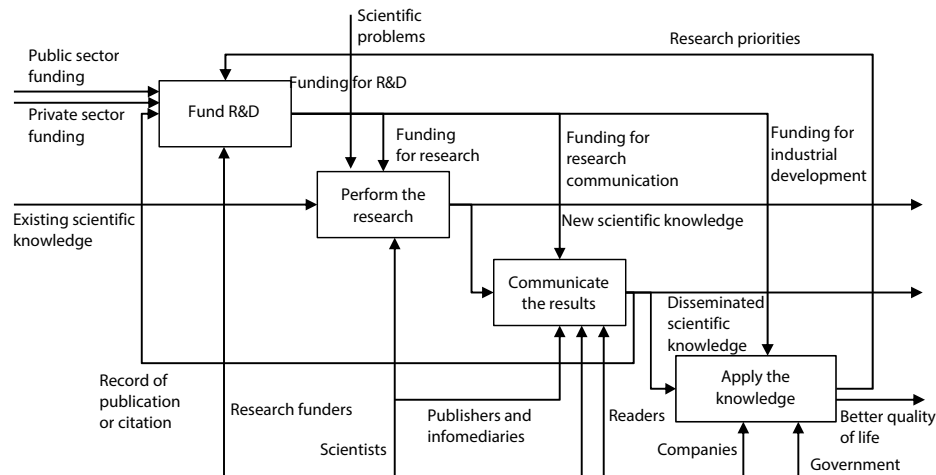
long and twisted, and pursued by various research groups from numerous angles. Parts of the path may be dead-ends (e.g. negative results), and different paths may lead to the same goal. On the way, new hypotheses are generated, and the potential outcomes can not be entirely predicted from the outset. In figure 4.1, Lewison's model (Lewison, 2004) of the process from funding to health improvement is illustrated, displaying a large number of the different elements which influence the process, and the complexity of the entire system is apparent. If we consider this complexity and the iterations which might be necessary for a given theory, idea or method to result in an actual treatment change, we can see that using "improved health" as an endpoint for measuring research quality, only a very small portion of medical research has a directly measurable quality. This only further illustrates the complexity of the research quality concept and the danger of attempting to define hard endpoints for an assessment thereof.

Figure 4.1: Model of interactions in medical research. Adapted from Lewison (2004).



As we have seen in the preceding chapters, evaluative bibliometrics as well as peer assessments, traditionally focus on research outputs such as journal articles, visible in the left part of fig. 4.1. While there is some circular interaction in this model, the general flow of events is from left to right, and despite these loops, or iterations, the basic idea is that funding of research through one way or another should lead to effects on health, however; the path leads through publishing of research results. Also in the more general model of scientific communication by Björk (2007), we can see the importance of research output, and especially the documents produced by researchers, in the communication process (see figure 4.2), and Bazeley (2010) also considers scholarly dissemination to be one of the pillars of research quality. When referring to research publications the common understanding is often that of journal articles. However, this

Figure 4.2: A model of scientific communication as a global distributed information system. Adapted from Björk (2007).



is not necessarily given, as different domains may weigh journal articles, conference papers, monographs, anthologies, data sets, software and other publication types differently. These weights are part of the scientific discourse of any domain, as are other publishing traditions, such as the length of reference lists, the typical age of references and the ordering and number of authors of a publication (e.g. Todorov & Glänzel, 1988; Baerlocher et al., 2007; Leydesdorff & Opthof, 2010b; Waltman, van Eck, van Leeuwen, & Visser, 2013).

As we can see from figure 4.1, the document types are here divided into papers in journals, papers at conferences, other documents and books, and also clinical guidelines. A few studies have looked at the latter as potentially useful for bibliometric studies (Lewison, 2002a, 2004; Andersen, 2013), and thus evaluate a different category in the model, than the traditional document types utilised in bibliometric evaluation. As mentioned above, the traditions which create the structure of the publishing system and the evaluative consequences are domain-dependant and it is essential to understand these traditions in order to gain an understanding of research quality in the domain. In the following section we will therefore describe medical research as a field. While several specialties, disciplines and research areas ranging from basic health science to applied and implementation research can be placed under the broad category of medical research, we will argue for the inclusion of each of these fields in this chapter.

4.1. A very brief history of medicine

Medicine, medical research, or the medical domain may be understood very differently depending on the reader's culture, religion or personal beliefs. In the context of this thesis, any reference to medicine is meant as contemporary, western medicine.

In order to clarify this, and to give the reader a taste of the setting, a short account on the history of western medicine will be given here. The focus of this section, however, is on a more thorough description of the current paradigm in western medicine and the implications for research assessments, scientific communication and publishing traditions.

Not unlike mathematics, physics and philosophy, modern medicine has a close connection to the work of ancient Greece, although older forms of practice have existed (Nutton, 2006). Some very visible manifestations of this are the Staff of Asclepius and the Hippocratic Oath (Mory, Mindell, & Bloom, 2000). The serpent-entwined staff of the Greek god of medicine and healing Asclepius is a commonly used symbol in the health-care environment, used to symbolise eg. clinics, ambulances and medical drugs. The Hippocratic Oath may be traced back to Hippocrates of Kos (ca. 460 BC – 370 BC), or possibly students of him (Markel, 2004). The oath has been used as an ethical code of conduct by practitioners of medicine in many countries for many centuries, albeit in modified, localised forms. The Hippocratic School perceived illness as an imbalance of the four humours (humourism); blood, black bile, yellow bile and phlegm, and a cure required restoring the balance between the humours – often achieved through venesection, or bloodletting. Quite often the focus of the practitioner was on prognosis rather than diagnosis (Garrison, 1966). Hippocrates is commonly referred to as the father of western medicine. His intellectual successor was the Roman-Greek surgeon Galen, or Aelius Galenus, (AD 129 – ca. 217) whose work remained among some of the most central until the 16th century. Galen authored roughly 600 medical treatises on various topics during his lifetime (Kotrc & Walters, 1979). However, one of his primary achievements was the documentation of the human anatomy in such detail it had never been seen before. It is in particular this part of his work which remained a cornerstone in medical teachings through the European middle-ages.

Early christian medical practice gave the world the first hospitals and the teachings of Galen were largely endorsed by the Church. However, during the Dark Ages, medical practice was largely controlled by the Catholic Church, with few practitioners, few translated texts and a practice which was still based on Galen and venesection (Nutton, 2006).

Around the 11th century the first steps were taken towards re-establishing the profession outside of the church, with the establishment of the *Schola Medica Salernitana*, the medical college of Salerno, but also through heavy influence from the Arabic world. In Arabia, medical practitioners learned from the East as well as the West and were not as prohibited by religion, as early Muslim practitioners were not as strictly controlled by the Qur'an as their christian counterparts by the church (Nutton, 2006). Among the most influential Arabic personae on European medicine was Al-Kindi (ca. 803-873, most famously known for *De Gradibus*), who was among the first to seriously apply mathematics and quantification to medicine and pharmacology (Prioreshi, 2002). Also Abu al-Razi (865-925) and Abu al-Qasim (936-1013) played important roles, the former through experiments proving false the theory of humourism (Meyerhof, 1935;

Amr & Tbakhi, 2007), and the latter through his exceptional work on surgery (Syed, 2002).

The real revolution of western medicine occurred in the early 16th century when two prominent figures substantially changed what had been believed for almost two millennia. Andreas Vesalius (1514-1564) is referred to as the father of modern anatomy, in particular with his book *De humani corporis fabrica* (published 1543), which by some is seen as one of the founding documents of the scientific revolution along with e.g. the astronomer Nicolaus Copernicus' *De revolutionibus orbium coelestium* (also 1543). Also William Harvey (1578-1657) played an important role in the initial modernisation of medicine by correctly describing the systemic circulation and introducing the scientific method to medicine. The arrival of the scientific revolution and in 1620 the modern scientific method (introduced by Francis Bacon in *Novum Organum*) entirely changed the practice of medicine and medical research – albeit with strong relationships to the traditions and knowledge of antiquity, e.g. in the sometimes mechanical view of the human body.

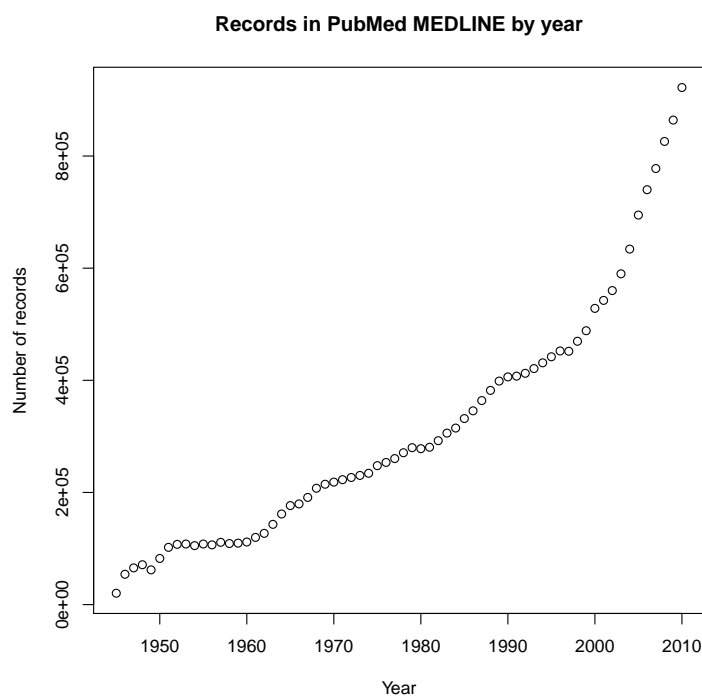
The first English medical journal, *Medicina Curiosa*, was published in 1684, followed by other prominent journals such as *Medical Repository* (1797), the *Philadelphia Journal of the Medical and Physical Sciences* (1820) and *The Provincial Medical and Surgical Journal* (1840, from 1857 known as the *British Medical Journal*) (Kahn & Kahn, 1997; Colman, 1999). The number of journals and papers published therein sped after the Second World War, and has increased ever since (for illustration see figure 4.3).

During the history of medicine, it has become increasingly scientific, systematic and research-intense, to a point where it is one of the most publishing research fields at all. As can be seen from the above, it is also a field with a well-documented history, and also featuring several meta-journals dedicated to the history of medicine, medical ethics, research methodology and method development. As such it is a field which is very likely to have a good understanding of itself.

4.2. Evidence-based medicine and scientific communication

Within the Western world, the evidence-based approach to medical practice and research is the dominant view, incorporating its own quality system. The following section will present the research paradigm as well as the research on some of the flaws, strengths and assumptions of evidence-based medicine (EBM) in addition to the relationship to bibliometrics and research quality. While evidence-based medicine (EBM) is only one of several approaches to medical and health scientific research, it has become the dominant one since its introduction in the early 1990's (Moskowitz & Bodenheimer, 2011). EBM integrates well with the reductionist paradigm present in health science, as opposed to the more humanistic, holistic approach which has gained further attention in the current debate, indicating a divergence between the two paradigms (Hunter, 2008; Miles, 2009). It is not the intention of this section to discuss or promote one paradigm over another; although it focuses on EBM, as this principle is al-

Figure 4.3: Records (all publication types) in the PubMed MEDLINE database published between 1945 and 2010, as of late 2011.



ready firmly established and widely accepted, while also relating to the research quality concept.

The concept of evidence-based medicine is based on the principle of creating and using the best possible evidence for the treatment of patients rather than relying purely on experience and assumptions. This evidence should be found in treatment trials following a strict, scientific approach and syntheses of such (meta-analyses). The approach requires medical practitioners to keep up with current research, perform literature searches and critically appraise the literature. EBM is a way of thinking research as much as practise, although there may be some problems in adapting research into the treatment of patients (Moskowitz & Bodenheimer, 2011). The concept is most clearly present in clinical research, that is research focusing on the treatment of patients, including prognosis, diagnosis, prevention and other aspects of treatment. Medical research also includes preclinical research, the medical equivalent to basic science, focusing on developing new theories, testing hypotheses on animals or performing biomedical experiments prior to clinical testing of e.g. drugs. More recently translational medicine has evolved as a postclinical research field, aiming at translating clinical findings into practice (Marincola, 2003). These research areas are also referred to as basic, applied and translational research, or biomedicine, clinical medicine and translational medicine. We will use the terms biomedical, clinical and translational research here. With regard to the evidence concept, it is present in each of these fields, but due to the experimental

nature of biomedical research, the perception of evidence may at times differ from that found in clinical research. This dissertation will include views from each research area despite these different understandings of the evidence concept, and the potentially different perceptions of research quality following from this. However, all three research areas are inescapably intertwined and should be understood in relation to each other.

To further elaborate on the EBM concept, the definitions of EBM are many, but a very precise one is stated by Greenhalgh:

“Evidence-based medicine is the use of mathematical estimates of the risk of benefit and harm, derived from high-quality research on population samples, to inform clinical decision-making in the diagnosis, investigation or management of individual patients”. (Greenhalgh, 2010, p. 1)

There are several important issues in this statement which together are defining for what EBM is. First of all, we can see how decisions for the individual are based on studies of population samples. While this to the layman may seem like a generalisation of the individual, and the degradation of diseases to numbers and figures, the use of evidence allows medical practitioners to make decisions based not just on what they can see in the single case, but also what is known from other cases. If for instance a patient suffers from chronic kidney disease, how large is the risk for a myocardial infarction? Should special attention be paid to patients with previous strokes? Through statistical (frequentist) inference the likelihood of the risk can be estimated, based on what has been observed (and documented) in larger samples than the single patient. The typical design for testing a treatment involves statistical significance testing of the *null-hypothesis*; that there is no observable difference between the treatment group and the control group (receiving placebo or traditional treatment) (Goodman, 1999; Shaw & Chase, 2007). There are some problems to this approach, in particular in their interpretation: an observed *P*-value under a given value (typically 0.05) indicates the probability that the null-hypothesis is true, however, it is often reversed so a probability of 5% for the null-hypothesis is interpreted as 95% odds that the null-hypothesis is false (Goodman, 2008). This conclusion can not be drawn, nor can conclusions about the effect of a treatment be drawn from such significance tests. A statistically significant result merely indicates that the sample size is large enough to exclude chance as the cause of observed differences between treatment and control groups, and that statistical significance should not be confused with clinical significance (Goodman, 2008). This does not mean that statistical tests of large samples cannot give us an indication about the likelihood of a risk or treatment success, and lead to protocols or guidelines that improve clinical decision-making (Morris & Ioannidis, 2013). This approach is ultimately utilitarianist in that the benefit of the many may come at the cost of the few (Shaw & Chase, 2007; Morris & Ioannidis, 2013). This follows from the fact that conclusions about the specific cannot be drawn from the general, we can only estimate, i.e. an array of symptoms may lead to a correct diagnosis 95% of the time, if diagnosis is based on clinical trials, however; 5% of the cases would be diagnosed falsely if no

other inference is used. Morris and Ioannidis emphasise the need for very large samples (megatrials) and the uniform application of methods, for comparison of results, if clinical decision-making should benefit from statistical analysis of treatments. Combined with the previous quote from Greenhalgh (2010), we can thus see how *evidence* plays a role of internal quality-rating system in medical research, that methods and sample sizes are related to perceived quality, and that the final endpoint is always clinical decision-making with the aim of estimating risk and benefit for patients.

With regard to the approach of evidence-based practice, it has been described as a five-step approach to finding the evidence for a specific problem (Sackett, Rosenberg, & Gray, 1996):

1. to convert our information needs into answerable questions (i.e. to formulate the problem);
2. to track down, with maximum efficiency, the best evidence with which to answer these questions - which may come from clinical examination, the diagnostic laboratory, the published literature or other sources;
3. to appraise the evidence critically (i.e. weigh it up) to assess its validity (closeness to the truth) and usefulness (clinical applicability);
4. to implement the results of this appraisal in our clinical practice;
5. to evaluate our performance.

As we can see from these five points, many sources are included in the process, including examination of the patient, but also expert opinions and other non-written sources may be considered. However, in the EBM-literature there is a strong focus on searching and assessing the research literature as the most extensive and reliable source of evidence.

Regarding the assessment of the research literature, several approaches have been made to classify studies based on their general design, as some study designs offer more (reliable) evidence to the reader than other. In general these approaches agree on the systematic review and meta-analysis to contain the most evidence, often followed by the randomised, controlled trial (RCT). These “levels of evidence” are currently represented most detailed in the Oxford Centre for Evidence-based Medicine (OCEBM) Levels of Evidence (OCEBM Levels of Evidence Working Group, 2011). The very distinguishing feature of this system is the classification into types of problems, and the correspondingly varying levels of evidence. Where simple evidence classifications typically only report levels for one problem type (therapy), the OCEBM distributes different study types across therapy, prevention, aetiology and harm; prognosis; diagnosis; differential diagnosis; and economic and decision analysis. For each of these problem types evidence is graded on a 9-point scale¹⁹. Common for all problem types is

¹⁹The scale is dual in that it can read as 1-4 and a more finely granulated 1a; 1b; 1c; 2a; 2b; 2c; 3a; 3b; 4

the placement of particularly well-executed systematic reviews on the highest evidence level, e.g. the best evidence for therapy studies should be found in systematic reviews of randomised, controlled trials, while the best evidence for diagnostic studies should be found in systematic reviews of inception cohort studies or clinical decision rules. Study types categorised on the other evidence levels vary more, according to the problem type, and the OCEBM is thus a very extensive approach to evidence-grading of study types. With regard to the above-mentioned connection between “high-quality research” and the inclusion of large sample sizes in specific methodological frameworks, the OCEBM levels of evidence are also a rating of a specific dimension of research quality. This dimension is concerned with the utility of research, but also the rigour and methodological strength, and might be used in bibliometric evaluation, if the study type can be extracted from the text.

Another well-established practice in medical research should be mentioned here, as it also has some influence on the scientific communication and conduction of medical research, namely the clinical research protocol. This ethical tool is applied in most western countries in order to ensure acceptable ethical conduction of research. The World Medical Association “Declaration of Helsinki” defines what are considered ethical principles for medical research on human subjects; this includes the research protocol, which must be submitted for ethical approval prior to conducting any research on humans. The research protocol outlines the research project in terms of hypothesis, methods, materials, patients, expected outcome and clinical importance (World Medical Association, 2002). Also the International Conference on Harmonisation (ICH) guidelines for Good Clinical Practice (GCP) builds on the Declaration of Helsinki (Informa Healthcare, 2005). The GCP guidelines define standards to be used by e.g. ethical committees when considering research protocols.

The protocol is not only an ethical threshold though, but a well-written protocol also serves as a ready article minus the results. It appears natural to assume this will also lead to a greater completion rate of started research papers, and it might even lead to a higher publication rate of negative findings – which would be a great advantage compared to many other research areas where the lack of reports on negative findings leads to redundant research. However, no conclusive research on this topic was found.

A very specific product of evidence-based medicine is clinical practice guidelines, which are intended to summarise the current evidence on a particular subject, and suggest to the practitioner how to perform certain tasks. There is however some debate on whether these guidelines should actually be trusted as what they claim to be, or if they are rather summaries of “what we think we know”. A number of studies have investigated this, e.g. Navarro Puerto et al. (2008) who validated 117 clinical practice guidelines against the AGREE²⁰ instrument. The AGREE instrument is used to assess the rigor of clinical practice guidelines, and it was found that 63% of the included guidelines were of poor quality. Certain factors influenced the quality of guidelines. In particular, national guidelines, those based on systematic reviews and those using ex-

²⁰Appraisal of Guidelines Research and Evaluation

PLICIT evidence-linked methods for deriving recommendations and those being regularly updated and featuring implementation strategies were more likely to be of high quality, than their counterparts. The AGREE instrument was also used in three other studies, analysing clinical practice guidelines in different cardiovascular diseases (Ferket et al., 2010; Ferket, Genders, et al., 2011; Ferket, Spronk, Colkesen, & Hunink, 2011). One study on cardiovascular risk assessment guidelines found very large differences in their recommendations (Ferket et al., 2010). They also found large differences in AGREE scores and only slightly more than half of the included guidelines were considered rigorous. Conflicting recommendations were also found for clinical practice guidelines on peripheral artery disease screening (Ferket, Spronk, et al., 2011) and imaging of asymptomatic coronary artery disease (Ferket, Genders, et al., 2011). Both studies found large variation in AGREE scores (33% to 81% (Ferket, Spronk, et al., 2011) and 21% to 93% (Ferket, Genders, et al., 2011)). Many guidelines were also created in collaboration with medical industry, which were not of the highest quality (Ferket, Genders, et al., 2011). In conclusion, while some clinical practice guidelines do as they intend and provide good clinical evidence, others do not, and naturally this leads to conflicting recommendations. It is not clear whether this is a general problem, or only for the observed areas, however; there is no reason to believe it should be limited to just these topics.

While clinical practice guidelines are intended to gather the evidence from other studies, these other studies are ordered into a very clear hierarchy of evidence-generating study types, as briefly mentioned above and further discussed in section 4.2. Different studies have found a connection between these study types and the citations they receive (e.g. Kjaergaard & Gluud, 2002; Patsopoulos, Analatos, & Ioannidis, 2005; Andersen & Schneider, 2011). The causality behind this mechanism, or whether it is pure association, has not been investigated further, but it could be speculated that certain study types are more clinically relevant and therefore are cited more.

To summarise the above, medicine is a millennia-old profession which has had an exceptional boom during the last century, in means of publications, personnel and economy. As a research field it has been clearly defined, and in some cases health science is even referred to as the fourth group of research fields (Science, Health Science, Social Science and the Humanities). Important standards and traditions for research and publishing have been established and formulated, and as a consequence it is quite likely one of the most popular fields for biblio- and scientometric analysis. It is therefore believed that an investigation of the research quality concept in this field will be particularly useful for the research communities of bibliometrics and medicine alike as well as general society.

5. Positioning the dissertation

Over the course of the previous chapters we have looked into the existing research on bibliometrics, focusing on the theoretical foundation for various bibliometric units and citations in particular. We have also positioned the dissertation epistemologically as well as ontologically, and thereby provided a framework for how research quality can be conceptualised, and dimensions hereof quantitated. In the present chapter, we will summarise the findings of the background and reviews with respect to the research quality concept. This will be followed by a brief recapture of the epistemological stance taken in this dissertation, in order to clarify how the research questions will be addressed before the specificities thereof are described in the next chapter.

We have seen that assessments of research can be divided into different forms of peer assessment and metrics-based evaluations. Each of these types can make judgments about different aspects of research quality, but the publication output of researchers is one of the most used, and the central unit of all bibliometric evaluation. While many believe peer assessments to be the gold standard for anything quality-related, we have also seen studies showing how metrics may provide additional insights, and that a combination may give us the best picture, especially as several flaws are pointed out for both approaches. With regard to identifying different dimensions of research quality from the literature, we have seen that the dissemination of research is one of other key dimensions in the research of Bazeley (2010). Refereeing is seen as a quality insurance tool in the dissemination process, with very different findings, ranging from “biased” and “flawed” to “useful” and “necessary”, but with some agreement that it should not be used as a quality assessment tool. The only peer assessment type which has been found useful for this, is peer ranking, e.g. of authors, journals or universities. This type of assessment may provide direct insights into the quality of research, however; depending on what the peer reviewer considers high quality, thereby being biased. As a contrast, citation analysis may provide more objective assessments, but only as a proxy of the internal quality of documents, and preferably as a statistical aggregate of several documents (e.g. the complete annual production of a university department). We have

also seen that bibliometric evaluation of research is highly complex, with several parameters and levels of analysis. This allows for a wide range of evaluation types, but also makes the interpretation of results difficult. As an example, measuring a temperature generally is a very low complexity task, and also easily interpretable by anyone familiar to the applied temperature scale - however, even when we know the specificities of a domain or subject area, citation analysis is difficult to interpret, as there may be numerous mechanisms behind the results. This is clearly illustrated by e.g. outliers, which are not merely measurement errors, but very real cases, that must be taken into account.

In evidence-based medical research there is an internal quality system (*evidence*), and a perception of some research designs as higher quality than other. It is possible that certain features of this quality system - as it is very publication-based - may be bibliometrically investigated, and some studies have pointed towards connections between citations and levels of evidence. An investigation of how medical researchers perceive research quality, the first research question of this dissertation, might reveal other dimensions of research quality, and by connecting these by their common features (the second research question), it is our aim to describe some of the mechanisms that affect citation analysis. Thereby we can achieve better interpretations of citation impact studies in medical research, or possibly develop different types of citation analysis. Creating this connection between the dimensions of research quality and bibliometric methods, is the aim of our third research question.

From an epistemological viewpoint, we use Bunge's philosophical approach to the world of sociology (Bunge, 1993, 1999), and accept that anything may be quantitated; that measurement is not implied in this; and that these quantitations are not absolute truths, but fallible steps towards a best representation of the truth. We acknowledge that research quality as a concept has elements of a social construct, as it is dependant - at least partially - on the social norms within a given domain. With regard to the research questions, this means that we will quantitate the different perceptions of research quality, gaining an impression of how well they are represented and agreed upon, but also that we will create quantitations of the final dimensions of research quality.

Our study design must support these assumptions and naturally provide responses to our research questions. Thus, we choose an approach incorporating qualitative research methods, to uncover the social aspects of research quality, with quantitative analysis of the latent structures behind the qualitative feedback. The exact approach will be detailed much further in the following chapter.

6. Methods and materials

The following chapter describes the methods employed to investigate the research questions of this dissertation. Initially, the overall study design will be outlined, in order to present to the reader an overview of the steps needed to reach the goal and thereby provide a link between the different methods.

The aim of the study design is to provide answers to the first two research questions of the dissertation, while the third question is answered through an analysis of the results from the other two. The study is thus focused on 1) collecting views on research quality from medical researchers and 2) quantifying these views on a larger sample in order to discover relationships between variables.

Following this description of the study design is a section on the applied methodology, and detailed methodological choices. As will be seen from the study design, an iterative approach is used. As a consequence, some data and preliminary results will be presented in these sections, to explain the concrete choices for the further method-selection and adjustment. The final results will be treated in the following chapter. The present chapter in its entirety should thus give an understanding of how these results were obtained, as well as why.

6.1. Study Design

To briefly recapture the objective of the study, we assume there is a connection between the concept *research quality* and *bibliometric measurement*. This connection could be described as a partial overlap, where not all of research quality can be described by bibliometric measurement, and not all bibliometric measurement is related to research quality. The connection between the two elements is closely connected to the logical flow from concept to metrics, illustrated on page 39, but with bidirectional logic:

concept ↔ qualities ↔ quantitation ↔ variables ↔ units ↔ metrics

, where the aforementioned connection is found between the left half, conceptualisation, and the right side, measurement. The right side is concerned with those topics treated in the previous chapters; bibliometric units and the variables or objects they are related to²¹. It follows from the above logic that a conceptualisation of research quality will give us insight into the connection between this concept and evaluative bibliometric methodology, which is the final goal of this dissertation.

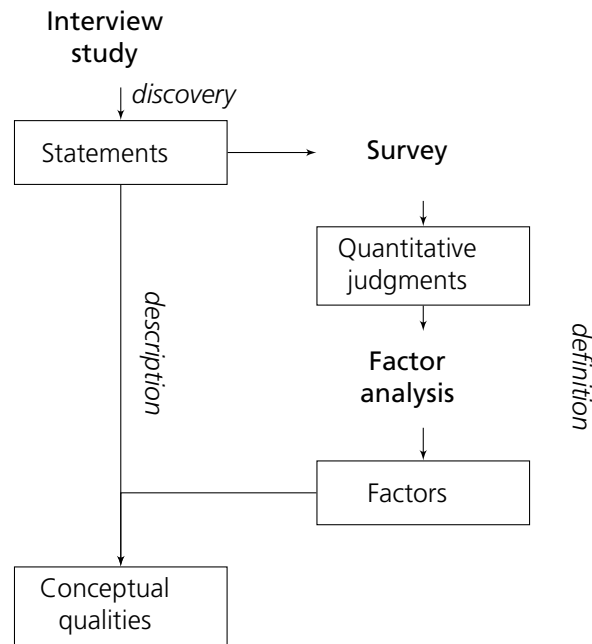
To achieve this goal, a mix of qualitative and quantitative methods will be used. At the heart of conceptualisation is a definition of the qualities of the concept; this was also discussed in section 3.1 and as argued there, we do not regard research quality from an essentialist perspective. Rather than trying to give a definitive definition of research quality dimensions, we aim at identifying a core of attributes, which together can be considered descriptive of research quality as a concept. This type of information is best retrieved through qualitative methods, enabling diverse responses from different viewpoints. While this type of data is correct and complete for some types of studies, it is also vulnerable to bias, and it is not likely to be generalisable. It is therefore considered desirable to create the definitions of qualities based on a quantitative verification of the qualitative findings. Such quantities also allow us to discover underlying structures and relationships between observations, by analysing co-occurring variations. The combination of qualitative and quantitative methods thus allows us to first discover potential qualities, quantitatively define these, create structure among qualities and finally by connecting the findings, using qualitative data to narrate the quantitative findings. This will be further elaborated in the following sections, but the above approach outlines the entire study on the general level.

Without going into too much detail at this point, a more concrete version of the above is outlined in figure 6.1, illustrating the steps and combinations of methods used to achieve the final goal. The initial stage involves interviewing medical researchers in order to gather qualitative statements on the subject, which are then validated through a quantitative survey. Using factor analysis it is possible to discover potential underlying structures in the resulting judgments, and by relating these factors to the original

²¹While metrics are the actual tools of bibliometric research assessment, they are not treated further here, as each metric is necessarily based on a unit and variable

statements, issued by the interviewees, we can expect to create qualitative narratives of relevant components of the research quality concept. These components should finally be discussed with respect to bibliometric measurement variables (not shown in the figure).

Figure 6.1: Model of study design



The initial steps of the study are somewhat inspired by a related study by Sternberg and Gordeeva (1996) which documented psychologists' conception of what made a research paper have an impact on the psychological field. Their study combined two approaches, namely initial interviews with twenty psychologists, asking them "to list characteristics of articles that make a high impact on the field of psychology. The psychologists were asked to list as many or as few items as they wished." (Sternberg & Gordeeva, 1996). This procedure resulted in a total of 45 descriptions, after the authors had removed redundancies. These descriptions were further qualified by asking a larger group of psychologists to rate these descriptions. This design can be seen as parallel to the interview and online survey used in this study, and their findings were used in the further design of the study. This is elaborated in the following sections, in which each of the elements in the study is further described.

6.2. Interview study

The interview study focused on generating statements on qualities or aspects of the research quality concept. As the goal is to describe and define research quality as perceived by the medical field, it appears prudent to include information from those

involved in medical research, either as producers (medical and biomedical researchers) or consumers (clinicians and general practitioners), henceforth known as users. User-involvement can be more or less controlled, which influences the type of data which can be gathered, and the research methodology to use. The choice of methodology is therefore dependent on the type of data which we intend to gather. The primary concern is whether data from users should narrowly represent their opinions and thoughts, thereby risking a great degree of spread in the data and potential lack of coherence in their type and scope, or if the gathered information should rather fit into pre-defined categories, with good coherence, but little room for creative or new answers. Both options are possible, and the interview can be weighted towards the open-ended or the coherent. Interviews conducted faithfully to a script containing primarily closed-type questions will result in short answers with good coherence, but very little room for discovery of unknown aspects (Kvale, 2009). These types of interviews may be used for confirming assumptions, but are hardly useful in the present case. Very open, loosely defined interviews will to a greater degree result in answers reflecting the opinions of the users, however with the risk of losing focus, as users may focus on less important aspects, or turn the interview towards other topics of their own interest. Other qualitative data acquisition methods, such as focus groups, may in some cases be considered superior to individual interviews in such regard. However, as focus groups include more participants, the total length of a session will naturally be longer than the individual interview. This was a concern, as an interview length of e.g. eight hours would certainly deter some potential interviewees from participating. A focus group approach may have provided a greater dynamic, and potentially additional data, however, as will be illustrated below, the interviews contained personal sections, ill-suited for a focus group. These considerations led to the omission of focus groups as a method. If the study relied more heavily on qualitative data, and in particular interpersonal interactions, the choice would have been different. As this is not the case, however, we will apply a semi-structured approach which should allow for creativity with open-ended and explorative queries, while also maintaining a script for keeping the conversation on track with relevant topics.

The previously mentioned study by Sternberg and Gordeeva (1996) is interesting, as it investigates something of a similar nature as research quality, namely scientific impact of research papers. While their object of study was different from the present, the concepts *impact* and *research quality* are so strongly related, they are occasionally used interchangeably, and their results could therefore be used as inspiration for what the interviewees in the present study might reply. It is also expectable that psychologists and medical researchers will agree on some aspects of research quality, and it should therefore be possible and relevant to use their responses as inspiration, although they are not a priori directly transferable. The results of Sternberg & Gordeeva can therefore be used as a background source for this phase of the study, providing key features for a semi-structured interview script. How these features were implemented will be described below, in combination with the exact approach.

Semi-structured interview script

The interview script was designed as a three-phase process, with the initial phase aiming at initiating dialogue and uncritically acquiring statements and thoughts related to research quality. The second phase was intended to qualify the first phase by discussing and elaborating presented thoughts as well as relating them to different perspectives on research and research quality. The final phase concluded the interview by asking users to associate to specific terms representing some of the key variables identified by Sternberg & Gordeeva as well as other related topics. The original interview script used by the investigator was in Danish language, as all interviews were conducted locally in a Danish hospital. A translated version of the script can be found in appendix C.

Initiating the dialogue was considered a very important part of the interview, as creating a setting of trust and openness between the interviewee and interviewer is very important for obtaining thoughts beyond the most obvious (Kvale, 2009). To help create this setting, users were asked prior to the interview to bring two or three research papers, which they considered to be of high quality. Each interview started by inviting the user to describe these papers and what they thought made them high quality research. In addition, each interviewee thus provided at least two references to research papers which they considered to be high quality. This gives us a collection of specifically selected high quality papers which can be compared to a background population with regard to any type of bibliometric measurement.

In order to ensure inclusion of relevant statements and observations from Sternberg and Gordeeva (1996), their main findings (see table 6.1) were summed into a number of key topics:

Confirmation Confirm or refute previous research, theories or hypotheses.

Interdisciplinarity Builds on interdisciplinarity or allows diffusion of knowledge or ideas between different areas of research, theoretical research and practice.

Integration Integrate results, theories and/or methods.

New knowledge Presents new knowledge.

New ideas Presents or generates new ideas, research or hypotheses.

Presentation Presents methods, results and implications in a logical and clear way. Also language correctness, flow and preciseness are included.

Relevance The research is relevant and has significance for either research or practice.

Validity and reproducibility Results are valid and reproducible.

These topics were included in the interview script, however, with no intention of forcing their appearance in the actual interview. If the course of the interview suggested it, the topics would be investigated. The topics should thus be seen as a type of checklist,

which should be investigated if mentioned by or hinted at by the interviewee. This was done during the entire course of the interviews, but focusing on the second phase, seeking to elaborate the initial dialogue and broadening the perspective.

Table 6.1: Basic statistics on item ratings for the Sternberg & Gordeeva (1996) impact questionnaire items. Means are calculated from ratings between 1 (low) and 6 (high).

Rank	Sternberg & Gordeeva items	Mean	SD
1	Makes an obvious contribution to psychological knowledge, adding something new and substantial	4.94	1.00
2	Presented results are of major theoretical significance	4.91	1.11
3	Presents a useful new theory or theoretical framework	4.79	1.18
4	Generates new research	4.70	1.15
5	Provides new and exciting ideas	4.57	1.18
6	Integrates many different areas of data previously thought to be unrelated or poorly explained	4.56	1.14
7	Integrates into a new, simpler framework data that had previously required a complex and possibly unwieldy framework	4.46	1.13
8	Opens up a new problem (research question) for investigation	4.45	1.08
9	Results of analysis are presented clearly and discussed carefully with tight, logical reasoning	4.43	1.25
10	The problem is clearly stated and well-conceptualized	4.33	1.33
11	Presents an unambiguous and creative interpretation of results	4.31	1.16
12	Topic is interesting and important	4.29	1.17
13	Presents a new viewpoint on a problem	4.28	1.03
14	Well-written, well-structured and well-organized	4.28	1.26
15	Provides a better explanation of existing phenomena	4.26	1.02
16	Contains useful implications for future research studies	4.25	1.10
17	Hypotheses are clearly stated and testable	4.14	1.30
18	The writing is succinct and internally consistent	4.13	1.31
19	Contains useful recommendations for changes or modifications in accepted theoretical constructs	4.10	1.09
20	Considers different alternative interpretations of the data and then arrives at a clear final, unambiguous interpretation	4.08	1.22
21	Has a logical flow and organization of ideas	4.05	1.29
22	Appears at the right moment, when people are ready to hear the message	4.03	1.54
23	Contains useful implications for scholarly understanding of the field	4.03	1.15
24	Contains some surprising results that make sense in some theoretical context	4.00	1.19
25	Contains interesting generalizations, which are clearly stated, confirmed, and based upon results	4.00	1.04
26	Contains useful recommendations for further research or for changing research methodology	3.98	1.10
27	Debunks an existing theory or way of thinking	3.88	1.39
28	Clarifies existing problem(s)	3.87	1.12

Continued on next page

Table 6.1 – continued from previous page

Rank	Sternberg & Gordeeva items	Mean	SD
29	Contains useful implications for theory building	3.84	1.22
30	Captures reader's interest	3.79	1.34
31	Results are of major practical significance	3.74	1.45
32	Starts and ends strongly, attracting attention and interest from the first paragraph and ending with a clear take-home message	3.72	1.37
33	Presents a new and useful test or technique	3.70	1.21
34	Demonstrates a useful experimental paradigm	3.68	1.13
35	Presents an experiment with a particularly clever paradigm or experimental manipulation	3.48	1.18
36	Is applicable to work in many areas of psychology	3.46	1.49
37	Provides evidence that fails to support an existing influential theory	3.42	1.29
38	Is clearly understandable to a broad cross-section of psychologists	3.40	1.39
39	Tone is unbiased and impartial	3.17	1.45
40	Presented findings are general ones	3.03	1.12
41	Results are intriguing but do not fit any existing theory	2.96	1.16
42	Presented theory is a general one	2.95	1.22
43	Contains useful information for professional practice	2.87	1.40
44	Provides evidence that supports an existing influential psychological theory	2.74	1.17
45	Includes concrete examples (from everyday life, lives of famous people, etc.)	2.32	1.26

As mentioned above, the final phase of the interview sought to elaborate very specific concepts and as such was more structured as the prior phases. Users were instructed to “brainstorm”, by responding with their immediate reaction to presented stimuli words. In this approach it slightly resembles the word association methodology, which is used mostly in psychology but has also been applied successfully in information retrieval and knowledge organisation research (Kiss, 1975; M. L. Nielsen & Ingwersen, 1999; Spiteri, 2002). In its essence, users are asked to respond to stimulus words using their immediate reactions, e.g. by saying/writing the first word they think of. This procedure may either be discrete (only one response per stimulus word) or continued (several responses to stimulus words), and it may be limited to particular response types (e.g. synonyms or pre-compiled answer choices). When comparing to this methodology, our approach is more open-ended by not limiting the response type and by not limiting the interviewee to single-word or single-concept responses. The stimuli words presented to the users were new knowledge, interdisciplinarity, impact²², citations, layout (presentation) and peer review. Some stimulus words were omitted in specific cases where the interviewee

²²Impact can be interpreted as importance, significance, relevance or the journal impact factor. The word is thus highly ambiguous, and no further explanation was provided. This was intentional, as the interpretation was considered just as interesting as the response. The other, related term *citations* was used to ensure a response to the citation-related aspects of research quality.

would bring the topics up independently.

6.2.1. Pilot interview study

A pilot study was conducted using the approach described above. The purpose of the pilot study was threefold:

1. To test the feasibility of the semi-structured interview as a method for retrieving statements or descriptions of research quality.
2. To provide feedback for the interview guide, improving the design for the actual study. Responses to the pilot study can be used as additional material for the final interview guide in supplement to the pilot study material.
3. To gain an impression of the required number of participants for the final study. It is expectable that the amount of unique answers will grow in a diminishing-returns sense (Weiss, 1994, p. 21), thus allowing identification of a good sample size.

Six users were invited to, and participated in the pilot study. The six users were specifically selected based on their profession, research activity and accessibility. Three of the participants were medical doctors, one was a biostatistician and two were nursing researchers. All six participants were active researchers, however, with varying research experience (ranging from low to moderately high). The medical doctors were chosen as representatives of the population sample to be used for the actual study, the other three participants were chosen to add outside, yet strongly related, perspectives to the study. Additionally the nursing researchers were accustomed to qualitative research methods, thereby potentially being able to give methodological feedback.

The participants received a letter with information about the study (see appendix B.1 for actual letter and B.2 for English translation). The information regarding the background of the study was very limited in order to create as little prior bias towards the topic as possible, while still providing enough information for the participants to relate. Each participant was asked to bring two or three research papers which they considered high quality, and they were also asked to sign a declaration of consent (see appendix C.3) and fill out a short form with details regarding their profession, educational background and research experience.

Each participant was provided with login-information for a booking system allowing them to book 30-minute slots at the most convenient time for them. All participants brought the requested materials to the interviews, which ranged in length from 20 minutes to 40 minutes. Interviews were conducted in Danish and transcripts can be seen in appendix A.1.

6.2.2. Script revision

The pilot study provided responses relevant to the topic of research quality, with diversity between the interviewees. The pilot study aimed at testing whether participants would communicate freely, or only respond in a closed manner; whether there was difference between the replies of different participants and whether their responses were at all related to findings from the literature as described in the previous chapters. All these criteria were met, and the approach was therefore considered to be feasible for the further data collection. The quality of the responses from the pilot study was so high that they might have been considered for inclusion in the final results, however, some minor changes to the interview script were made, and non-medical researchers participated. Hence, for the sake of comparability, data from the pilot study were not included after all.

The experiences from the pilot study led to minor changes in the final version of the interview script. The first phase, discussing papers brought along by the interviewees, was found to fulfill its purpose, as all users in the pilot study initiated dialogue based on this task, and provided initial input on their own thoughts on research quality. This section of the interview script therefore remained unchanged. For the following section, a few additions were made, based on statements put forth during the pilot study; questions regarding the connection between evidence and research quality, a potential food-chain from case studies to clinical practice guidelines and whether the participant was aware of the citations received by the abovementioned papers. While the other questions were used in liberal order, and omitted in some cases, the final question about citations was always asked towards the end of the interview, as it might have influenced the direction of the interviewees thoughts on research quality. This particular, factual query was primarily included to document whether the paper selection might have been influenced by citations to the papers. The final section of the interview script remained unchanged, although the order of presenting stimulus words was randomised.

The diversity of responses from the pilot study gave the impression that more participants would be necessary for the main study, as the final pilot participant provided substantially new responses. Through heuristic judgment it was approximated that 20 participants would be sufficient to cover common perceptions of research quality as well as a great deal of personal variations. An empirical test was conducted subsequent to the actual interviews in order to confirm this heuristic (see next section).

6.2.3. Main interview study

For the main interview study, all doctors employed at Aalborg University Hospital were contacted through email, with an invitation to participate in the study. 25 of the recipients replied positively and in turn received a letter with information on the study and booking system, a consent form and background information similar to that described in the above pilot study (see appendices B and C). The booking system offered potential participants to select from a wide range of time slots and dates. 14 of

the invited participants registered for a time slot. After the conclusion of these interviews, it was estimated that the inclusion of additional participants would not provide substantially new insights, and contact to the remaining eleven unregistered users was postponed until the initial data processing was concluded.

At the beginning of each interview the participants were informed that the dialogue would be recorded using a dictaphone and the articles they had brought were labeled for future reference. Each participant was assigned a label to distinguish them from each other using non-personal information. Using the articles to initiate dialogue was successful in every case, as participants freely talked about what they considered quality in those papers, and what made them stand out enough to select those papers. The interviewer sought to mostly ask open-ended questions, furthering elaboration of the interviewees responses and allowing conversation to flow freely. While the participants were informed that the expected time consumption was 30 minutes, this was not enforced, and the interviews lasted from 19 to 48 minutes, with an average length of 33 minutes. Due to this rather limited time, however, the interviews were somewhat guided towards specific topics (see the above description of the semi-structured interview guide).

All interviews were transcribed manually, using Express Scribe by NCH Software to manage audio playback. The transcripts (in Danish) can be found in appendix A.2. As can be seen, the transcripts were initially coded as questions (from the interviewer) and responses (from the interviewee). The transcription was performed solely by the interviewer, and kept as close to the spoken word as possible. In some rare cases, illegible words were omitted, and minor rephrasings were made in cases where the meaning would otherwise be unclear. Omissions were only made in cases, where words were completely illegible due to e.g. noise, and replacements were only made in cases where the transcriber was completely certain of the intended meaning. In any other case, the transcription was made exactly as spoken. All omissions were noted in the transcriptions.

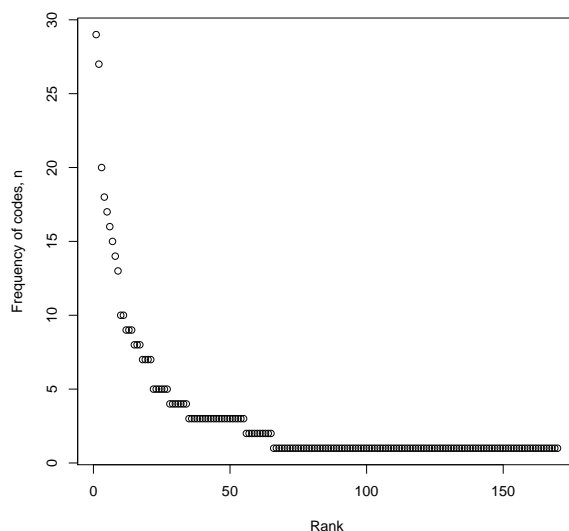
To further process the transcriptions into standardised, codified statements, the transcriptions were transferred to a system for manual analysis. The transcriptions were typeset using \LaTeX , which allows easy data conversion to other systems, as the files contained only raw text and two custom defined \LaTeX commands to discern between questions and responses. This advantage was used to subsequently transfer the transcripts to a MySQL database, in order to codify statements from the interviews. The database included not only the text, but also labels for the interviewee and information on the order of questions and responses. Using PHP and HTML a locally hosted web service was created. The purpose of this service was to present database entries in an easily readable manner and attach codes to the presented entries, linking these together. The codes were created by the same person as the other parts of this study. Upon entering text, suggestions for previously entered codes were shown (based on a dynamic AJAX query to the database containing the previously entered codes), allowing a certain degree of stringency. However, the initial codes were created with further distillation

in mind, and overlapping codes were allowed. Using the abovementioned, registered order among questions and responses, it was possible to present individual responses in a randomised order, while maintaining relevant context (current and previous question and response). By presenting the answers in random order, fatigue and contextual bias were minimised. Codes were created for each response, creating a link between the codes and the responses. A total of 170 codes were created based on the responses of interviewees (see appendix D). There was large variation in the frequency of the codes (see figure 6.2), and some overlap was clearly visible. The codes were categorised according to their content, initially revealing 22 groups:

1. Ethics in research
2. Factors related to the author(s)
3. Factors related to the impact of articles
4. Clinical practice guidelines as specific document type
5. Other specific study types
6. Groundbreakingness of research
7. Citations
8. References
9. Evidence
10. Research as a source of inspiration
11. Interdisciplinarity
12. Methodology and materials
13. Factors related to journals
14. Novelty
15. Results
16. Style (e.g. visual/textual)
17. Basic and clinical research
18. Implementation / translation
19. Peer review
20. Research environment
21. Clinical relevance
22. Various connections between other factors

In appendix D the codes are also divided into these classes. The classification was created manually, as a topical classification. The classification does not have an impact on the interpretation of the codes, but was used as a first step on the way to creating the

Figure 6.2: Frequency of response codes ordered by rank of occurrence



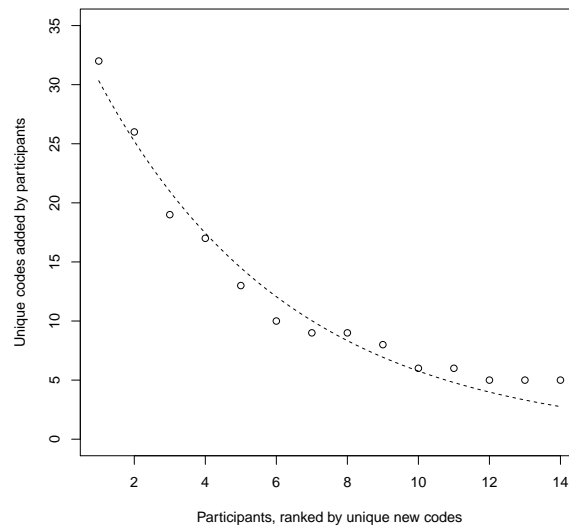
final statements for the survey and discover overlapping codes. We will therefore not discuss the principles for classifying the codes further.

Regarding the number of new, unique codes added for each participant, one would expect some exponential decay: As more interviews are added, the number of new codes diminishes. This is a natural phenomenon, also observed in e.g. usability studies (J. Nielsen & Molich, 1990; J. Nielsen, 1992), and thus allows an estimate of when it can be considered unfeasible to add more interviews. Figure 6.3 shows how this is indeed the case, when interviewees are ordered by the number of unique new codes provided by their responses.

While the addition of new interviews would certainly add something, it is estimated from the curve in figure 6.3 that this would hardly be justifiable from an ethical viewpoint, as the gain would presumably be diminutive, and no longer match the time invested by participants. The aforementioned remaining eleven unregistered users were thus not contacted to follow up on the lacking registration.

To recapture; through an initial pilot study, an interview study, including 14 participants, and a manual codification of interview transcripts, it was possible to create a number of statements related to the concept research quality. The further processing of the codes into human readable statements will be treated in the following section, which will also be concerned with quantifying the statements using an online survey.

Figure 6.3: Number of unique new codes added by each participant, ranked decreasingly



6.3. Online survey

The purpose of the online survey was to quantify the abovementioned statements in order to discover whether there are underlying structures in these statements. The statements were treated as dependent variables, and the co-variance of variables reveals whether there is any connection between variables. This process will be described in the following section on factor analysis. The demographic variables will be described towards the end of this section, and an initial analysis will be performed, with the purpose of assessing the representativeness of the sample. The actual analysis of variables depending on demographic parameters will be performed in the results chapter. The current section will focus on the process behind creating the survey, testing it and contacting participants.

A total of 170 statement codes were created from the interview transcripts, and we have seen how they varied greatly in frequency, several being issued only once, and with significant overlap between them. To minimise the number of codes, they were combined, and rephrased into questions or statements designed to be understandable without prior information. These statements are shown in appendix E, with a total of 105 questions, of which three are of a demographic nature (added as additional questions), 63 relate to different research procedures, norms and communication procedures, while the remaining 39 relate to article-specific factors, such as the visual style, methodology and the presence of specific sections. The overall division in these three groups of questions was retained in the final version of the survey form, however, considerable changes were made to the contents of each group. This was done over two main iterations, as the first version was submitted to a pilot test. There were no major design differences

between the pilot version and the final version. Rather than displaying both forms, only the final version is shown in appendix F, and changes are presented here. The pilot test was performed by presenting the survey to six participants, while the surveyor observed them. The participants were encouraged to “think out loud” while filling out the survey. If participants were hesitant at particular questions, the surveyor followed up on these after conclusion of the survey. Two of the participants were medical students (9th and 11th semester respectively), the remaining four had medical degrees. One of these was employed as the equivalent of a medical internship, and the three other were PhD candidates.

The demographic section was expanded to contain eight variables, aiming at elaborating on research experience, by including an estimate of the respondents publication count. Also questions regarding specialties, job function and work places were added. Several definitions of medical specialties exist, and the same is the case for job titles etc. The initial form only contained one field for job function, which led to some confusion among respondents. After consulting with a senior medical researcher and medical doctor, this variable was divided into two, one focusing on the primary job function (research, clinical practice, administration, management and other) and the other on experience, expressed as years since completion of formal medical education.

The second section, containing statements on research quality, was narrowed down to 32 variables, presented as statements to which the participants were asked to declare an agreement rate. These rates varied between two scales, a 5- and an 11-point scale. The scales were designed to have an absolute disagreement at 0, complete agreement at maximum value, and a midpoint at the median value. While the scales are in principle ordinal, these three reference points and the simulated equidistance between values allows us to perform operations on the degree of being “in” or “out” (equivalent to agreeing or not agreeing) (Ragin, 2008). As the final purpose of these variables are to identify factors, which influence the variables, we can use these values to create fuzzy sets of variables (Ragin, 2000, 2008) which together describe a more general topic. The scales purposely switched between 5 and 11 points for two reasons: 11-point scales were used in cases, where the statements reflected a question of belief, personal opinion or trust, while the 5-point scales were more concrete, although also reflecting personal opinions. Secondly, by using varying scales, the participants were indirectly forced to reflect more on their responses, rather than repeating the same value (fatigue bias). Several scales can be observed in the psychometric literature, and Dawes (2002, 2008) has experimentally tested the difference between using five and eleven point scales (Dawes, 2002) and five, seven and ten point scales (Dawes, 2008). In both studies a small variation in mean scores for the same questions was observed. This can likely be ascribed to data conversion procedures, as the scales used different values for minimum scores (0 and 1) in the first study, and a non-centered scale (1 – 10) in the second study. This assumption is further supported by very low differences in skewness, variance and kurtosis for either scale set. The advantage of minimising fatigue bias thus seems to outweigh the potential problems of using varying scales.

In this second section of the survey only minor rephrasings were made as a result of the pilot test; some statements were difficult to understand by the participants and were rephrased to be more clear. The original meaning of the statements remained unchanged. A few additional scale-labels were added to clarify their relationship to specific statements.

Compared to the initial survey form (appendix E), the third section was completely redesigned. In the final survey, participants were asked to select the three elements from each of the groups “Materials”, “Methods”, “Results”, “Style” and “Other”, which they would consider most decisive for the quality of an article. The original design asked participants to rate every element on a 5-point scale, however, it appears likely that e.g. statements concerning the method-section of an article might very likely be considered very important by all participants. Thus it would hardly be possible to tell anything about these variables. By selecting three elements in descending order, the participants are forced to decide what is most important. Each approach has its advantages and drawbacks, but as the more qualitative statements are treated in the previous section, and the elements in the third section appear to be of a more binary nature, the chosen approach was considered more feasible. All response options for the five topics in this section are documented at the end of appendix E. Variable names, wording of questions and response type is listed in table 6.2

Table 6.2: List of variables, question wording, response types and data types in the online survey.

Variable name	Question	Response type	Data type
Demographic variables			
birthyear	Year of birth	Drop-down	ratio
gender	Gender	Drop-down	nominal
experience	Years since completed formal medical education	Text field	ratio
jobfunction	Primary job function	Drop-down	nominal
specialty	Specialty	Drop-down	nominal
workplace	Primary work place	Drop-down	nominal
pubcount	Approximate number of publications where you are (one of) the author(s)	Text field	ratio
country	Country of residence	Drop-down	nominal
Main variables			
authorqual	Please indicate to what degree you believe the author(s) of an article to be decisive for how you experience the quality of an article	11-point scale	ordinal
authorexpr	Articles written by experienced authors are generally of higher quality	5-point scale	ordinal

Continued on next page

Table 6.2 – continued from previous page

Variable name	Question	Response type	Data type
goodauthor	If I know an author as a good researcher, I am inclined to perceive their new articles to be of high quality before reading them	5-point scale	ordinal
morejournals	"I find there are more research journals than necessary", do you agree with this statement?	11-point scale	ordinal
<i>Many journals limit the length of articles. Do you believe this to be:</i>			
limitproblem	- A problem, if the methods and discussion sections cannot be elaborated	5-point scale	ordinal
limitgood	- A good thing, as there is little time for reading lengthy articles, and they must therefore be short and concise	5-point scale	ordinal
editorial	If an article is mentioned in the editorial of a journal, it stands out from the other articles in that issue.	5-point scale	ordinal
journalquality	Some journals generally publish articles of higher quality than other journals.	11-point scale	ordinal
datasource	Please indicate on a scale whether your primary source for keeping up to date on new research is journals (e.g. subscriptions, e-alerts, table of contents, RSS) or by searching databases (e.g. PubMed, Embase)	11-point scale	ordinal
jifqual	The journal impact factor is indicative of the quality of articles appearing in a journal	11-point scale	ordinal
jifincent	High impact factors are an incentive when I select a journal to publish a paper in	5-point scale	ordinal
selref	To what degree do you agree with this statement: "When selecting references for my own papers, I am more likely to select ones from high-impact journals"	11-point scale	ordinal
qualindep	The quality of an article is independent of the journal it appears in	11-point scale	ordinal

Continued on next page

Table 6.2 – continued from previous page

Variable name	Question	Response type	Data type
guidequal	To what degree do you agree with this statement: "Clinical guidelines reflect the highest quality research available on a topic"	11-point scale	ordinal
guideref	References used in clinical guidelines are usually of especially high quality	5-point scale	ordinal
guideimplem	Clinical guidelines are decisive for implementing research in practice	5-point scale	ordinal
guidethink	Clinical guidelines are based on what we think we know, rather than evidence	5-point scale	ordinal
<i>If an article is cited by many other articles, I would consider this to indicate:</i>			
citequal	- High quality research	11-point scale	ordinal
citerespect	- Recognition / collegial respect	11-point scale	ordinal
citetrend	- The article is on a hot topic	11-point scale	ordinal
citeref	I am more likely to use highly cited articles as references	5-point scale	ordinal
citeping	I am more likely to cite the work of a well-known researcher in my field	5-point scale	ordinal
citecoll	I am more likely to cite the work of my colleagues	5-point scale	ordinal
groundbreak	Some research can be considered ground-breaking, which is an outstanding quality	11-point scale	ordinal
clinbas	Changes in clinical practice can be led back to advances in basic research	11-point scale	ordinal
baspurp	The ultimate purpose of basic research is improved health	11-point scale	ordinal
clinrel	Research - whether basic or clinical - should be clinically relevant	5-point scale	ordinal
gatekeeper	Peer review is a gatekeeper, assuring a minimum of quality in published research	5-point scale	ordinal
strictrev	More strict peer review would lead to higher quality in published research	5-point scale	ordinal
innovatebias	Peer review may prevent new and innovative or ground-breaking ideas from being published	5-point scale	ordinal

Continued on next page

Table 6.2 – continued from previous page

Variable name	Question	Response type	Data type
negfind	Negative findings may be interesting and important	5-point scale	ordinal
goldstand	Certain methods are golden standards and should be used whenever possible	5-point scale	ordinal

In order to progress through the survey, participants needed to provide a response to every statement. If a response was missed, by accident or purpose, the web page was refreshed and the missing responses were clearly marked and the participant was asked to complete the missing responses. This approach ensured that the final number of participants had provided responses to all variables.

One common problem of online surveys is a very low response-rate, however, as the aim of this part of the investigation is to do factor analysis, and not to generalise findings to the entire population of medical researchers, we were not interested in achieving a particular response rate, but a sufficient amount of responses. There is no clear definition of what a “sufficient” sample size for factor analysis is, but Gorsuch (1983, p. 332) suggests at least 100, but also at least five observations per variable. Other suggestions have been made, recommending at least 300, 500 or ratios of 3, 5 or 10 observations per variable MacCallum, Widaman, Zhang, and Hong (1999). As there are 32 variables to be used for the factor analysis, it is estimated that a sample size between 160 and 320 (sample-to-variable ratio between 5 : 1 and 10 : 1) should be sufficient. It is desirable to acquire the highest possible number of observations, as “larger samples will usually lead to clearer indications of the number of factors...” (Gorsuch, 1983, p. 147), but as with the number of respondents in the interview study, there is an unethical aspect in collecting more samples than necessary.

With the expected low response-rate three approaches were considered for contacting medical doctors, with priority given to the more exact approaches. The first approach was to contact medical associations, followed by university and teaching hospitals. If neither of these would give any results, the final approach would be to contact researchers directly based on correspondence-addresses in articles. All three approaches were limited to the EU-27, Norway, Iceland, Switzerland, USA, Canada and Australia. The initial contact to the professional associations yielded limited results, as only the Icelandic Medical Association was positive towards contacting their members directly, and the Dutch and British Medical Associations offered a notice in their member newsletters. While this approach might have provided a sufficient amount of observations, it was considered to be subject to selection bias, as only medical doctors from three countries would be included. Thus, the second approach was initiated, and contact to potential participants was planned to go through central hospital administrations. Contact information for all university and teaching hospitals within the stated nation-limits were collected from the online homepages of these. Far from all hospitals were possi-

ble to contact this way, and in those cases where it was possible, the communication would often go through public relations or press departments. The hospitals were on general not willing to forward an external survey to their employees, either with regard to ethical considerations or time expense, and only one of the contacted hospitals was clearly willing to participate. This was also considered infeasible and biased, and the final approach was used instead.

200,000 articles from all health-science categories in Web of Science were obtained through manual downloads. The articles were selected as the first 100,000 articles in 2011 and the same in 2012. From these articles, 173,211 e-mail addresses from corresponding authors were extracted, of which 36,855 were from countries outside the defined area, an additional 28,507 were from generic domains²³. The final sample of e-mail addresses was further limited to only contain addresses occurring in at least two articles, resulting in a total of 17,147 contacts. An e-mail was sent to each of these addresses, inviting the recipient to participate in a “Research quality survey”. A link was given to the survey, and the recipient was guaranteed full anonymity (see appendix G for exact wording). To spark the interest of the recipient, a reference and citation count for their most cited paper from the above sample was listed. Of these addresses, 1,081 were not delivered²⁴. From the remaining 16,066 addresses, 387 entered their e-mail address into the survey and 279 completed the survey. It is unknown why no further of the invited contacts accessed the online survey, but it is likely that an automatically sent e-mail, as the one used here, was caught by Spam-filters in several cases. As the number of recipients was within the desired range of sample size, however, no further e-mail addresses were contacted. For comparison, Sternberg and Gordeeva (1996) collected 252 responses from their survey. An overview of the demographic data for the final participants is shown in section 6.5.

6.4. Factor analysis

Factor analysis, in general, is a set of methods with the common aim of categorising a group of variables so that a number of qualitatively distinguishable factors can be identified from these variables, based on the degree of generalisability between these (Gorsuch, 1983, p. 3). That is also the overall aim of using factor analysis to distill the results from the online survey.

Several approaches for factor analysis exist, as the term covers a number of methods used for various purposes. The most common distinction is between exploratory and confirmatory factor analysis, but also different factor analytical models exist. This dissertation primarily uses Gorsuch (1983) as a source for information on these models and factor analysis in general. The selection of one model over another is dependent of the data, the distribution hereof and the purpose of the factor analysis. In this disserta-

²³e.g. @gmail.com or @hotmail.com

²⁴This covers e-mails where the recipient was no longer employed at the institution, faulty addresses, out-of-office replies beyond the planned time limit and automatic messages about failed delivery.

tion factor analysis was used to discover factors without rigid assumptions about their number or construction. In other words, we used exploratory factor analysis, and the chosen model should match this.

For any factor analysis, we require a matrix, X_{nv} , containing data for n observations of v variables, i.e. the observed variables in our sample. While this matrix consists of measured values, it may also be represented as a number of factors and variable weights, resulting in the same matrix, but with new information about related variables. The most straightforward representation is the full component model (eq. 6.1),

$$X_{iv} = w_{v1}F_{1i} + w_{v2}F_{2i} + w_{v3}F_{3i} + \dots + w_{vf}F_{fi} \quad (6.1)$$

, where w_{vf} is the weight of variable v on factor f and F_{fi} are observation i 's score on all f factors (Gorsuch, 1983, p. 20-21). This model will result in the same amount of factors as variables, but with different loadings (factor scores), allowing discovery of related variables from high factor loadings on the same factor. The common factor model approaches the subject differently by splitting factors into those factors which several variables have in common, and a unique factor for each variable. The model is defined by equation 6.2,

$$X_{iv} = w_{v1}F_{1i} + \dots + w_{vf}F_{fi} + w_{vu}U_{iv} \quad (6.2)$$

in which the weights, w_{vu} are the weights of variable v 's unique factor, and U_{iv} is observation i 's unique factor score for variable v (Gorsuch, 1983, p. 26). Both models may allow correlated factors or be restricted to uncorrelated factors only. The choice of allowing correlation or not is related to the *rotation* method used, which will be described further below. Generally uncorrelated factors are considered stronger, however; there is no theoretical basis for a priori assumptions of uncorrelatedness between factors, and we should therefore test it. This can be achieved by allowing correlation in an initial factoring test and subsequently adjusting the rotation according to the pres- or absence of correlation in this test.

As the output of the common factor model is what we wish to find, namely a set of factors which some variables have in common, this model was chosen for analysing the survey data. A number of relevant statistics can be obtained from the correlated common factor model, here taken from Gorsuch (1983, p. 29-33) and summarised as the communality of a variable, h^2 (Eq. 6.3), the uniqueness of a variable, u^2 (Eq. 6.4), the specific factor's proportionate contribution to the variance, s^2 (Eq. 6.5), and the error contribution, e^2 (Eq. 6.6).

$$h^2 = \frac{\sum w_k^2 + \sum \sum w_j w_k r_{jk}}{V_X}, j \neq k \quad (6.3)$$

$$u^2 = 1 - h^2 \quad (6.4)$$

$$s^2 = r_{Xt}^2 - h^2 \quad (6.5)$$

$$e^2 = 1 - r_{Xt}^2 \quad (6.6)$$

For equations 6.3 to 6.6 X is a variable in the observed set of variables, V_X is the variance matrix for that variable, r_{jk} is the correlation between variable j and k , while all w_i are weights as used above, and r_{Xt}^2 is the true score variance. The communality is the proportion of a variables variance which is described by the common factors.

As stated above, there are no rigid assumptions about the number of expected factors, however, the common factor model requires a prior number of factors. This number can either be found manually, based on theoretical assumptions or other empirical findings, or it can be found semi-automatically, using Cattell's scree-test (Gorsuch, 1983). This approach for empirical induction of an appropriate amount of factors is based on a subjective, graphical assessment of the eigenvalues of the factors. We will use this approach, however, with the addition of the optimal coordinates approach of Raïche, Walls, Magis, Riopel, and Blais (2013), adding an objective layer that has been found to provide very good predictions of the correct number of factors. The determination of whether the number is correct is a question of testing whether the identified factors can be considered statistically significant. Bartlett's significance test is very useful for this purpose (Gorsuch, 1983), and will be applied to factor analytical results to determine whether the identified number of factors is an expression of chance, or if the sample size is large enough to support the factor constructions.

Exploratory factor analysis uses the concept of rotation to further simplify findings and make the structure between factors clearer. Two main approaches for rotation exist, orthogonal and oblique, corresponding to uncorrelated and correlated factor analytical models respectively. Several rotation procedures exist for both of these, Gorsuch lists 15 oblique and 4 orthogonal rotation procedures. The varimax procedure is considered to be the primary orthogonal rotation approach, as it maximises the variance of squared loadings of a factor. For oblique rotation, the choice of procedure depends on e.g. the desire for high or low correlation. We will be using promax, an oblique procedure based on orthogonal rotation, and able to create a solution with high factor loadings from a generalisation of any orthogonal rotation.

The final element to be considered for the factor analysis is the treatment of variable scores. As the main variables are scored on both 5- and 11-point scales, direct comparison is not straightforward. As mentioned previously the scales do operate with clearly defined end- and midpoints, enabling some comparison between them. Other methods for making different (ordinal-like) scales comparable include calibrating categories to index values or standardising scores. We will apply both index values and standardised scores to the raw scores before performing the factor analysis. The two approaches will then be compared to gain an impression of the robustness of the test. It is desirable that the two calibrated data sets should create the same number of factors, and that common factors should have high loadings on the same variables. We will therefore first use the optimal coordinates approach on both data sets, to find the number of factors, and subsequently keep this number stable while testing for loadings, thus only changing one

parameter between the two tests. The raw scores for observation i on variable v are given as $\{v_i | v_i \in \mathbb{Z}, 0 \leq v_i \leq (M - 1)\}$ for M -point scales. The index value of observation i on variable v , I_{iv} , will be created as in eq. 6.7 while the matching standard score, z_{iv} , is given in eq. 6.8;

$$I_{iv} = \frac{100 * n_{iv}}{M_v - 1} \quad (6.7)$$

$$z_{iv} = \frac{n_{iv} - \bar{n}_v}{S(n_v)} \quad (6.8)$$

, where n_{iv} is the raw score of observation i on variable v , M_v the number of categories on the scale of v , \bar{n}_v is the average of the n_v distribution and $S(n_v)$ the corresponding standard deviation. The values of I are expressed as values from 0 to 100 with a midpoint of 50 as a consequence of the set of possible values for v_i (see above) and the formulation of I_{iv} . The values of z are given as products of the standard deviation, thus $z_{iv} = 0$ is equivalent to $v_i = \bar{n}_v$. Both distributions thus have specific midpoints, however; with very different meanings. As factor analysis is based on differences in variance, this should have limited consequence for the factor loadings, but the subsequent interpretation of the two calibrations is very different.

To sum up, following a robustness test of variable calibrations, we will be performing exploratory factor analysis, using the common factor model with promax (oblique) rotation. If no correlation is identified between factors, the factor analysis will be repeated with varimax (orthogonal) rotation. The number of factors is identified using the optimal coordinates procedure and evaluated using Bartlett's significance test. If the identified number of factors is not statistically significant, the factor analysis will be repeated until the largest, statistically significant number of factors is found. For the purpose of factor analysis, we use the `factanal` command in R (version 2.15.2 x64) (R Development Core Team, 2008), and the `nFactors` package (Raïche, 2010) is used to perform optimal coordinates tests. Bartlett's significance test can be incorporated directly in the `factanal` command, as can the choice of varimax and promax rotation.

6.5. Demographic analysis

Below follows a description and analysis of the demographic data obtained from the online survey, with the purpose of describing the sample and its representativity, discovering potential, systematic flaws and analysing the demographic variables for any latent structures which might influence the remaining results. The first part of this section thus summarises the demographic variables, followed by a more in-depth analysis, including factor analysis of the demographic variables. It is not given that any demographic factors exist, however, if they do, we must take these into account when analysing the further findings.

6.5.1. Demographic variables summarised

As mentioned previously, the survey contained eight demographic variables, which are listed in the initial part of table 6.2.

Year of birth, y , was modified as $\text{age} = (2012 - y)$. The experience and pubcount variables were registered through text fields, allowing user input, while all other variables only allowed predefined values through drop-down selectors. As might be anticipated, some of the entered values contained extra information, and were modified as corrected values where this was obviously possible, and classified as missing values (N.A.) in all other cases (see appendix H for documentation).

The distribution of gender is seen in table 6.3, and while more than twice as many of the participants were male ($n = 195$, female $n = 84$), this is not surprising. As can be seen from figure 6.5, a very large proportion of the included participants are from the United States of America, where the overall distribution of gender among physicians was 28% female as of November 2012²⁵.

Table 6.3: Distribution of gender

Gender	n	%
Female	84	30.11
Male	195	69.89

The age distribution shows a normal trend (see table 6.4) which is not reflected in the corresponding experience. This is due to a number of participants explicitly commenting (either in the experience data field or in the final comments field) that despite being medical researchers, they are not medical doctors. Based on these replies the wording of the question can be considered lacking, as e.g. microbiologists, hospital physicists or midwives can be considered as medical researchers as well as medical doctors. We can thus expect some error in this field, however, judging from the correlation between the two variables (see figure 6.4) there is a medium high correlation coefficient ($\rho = .78$) despite the increased presence of 0-values.

Clearly the majority of participants ($n = 258[92.5\%]$) were primarily employed in research or clinical positions (see table 6.5). Given the contact procedure, this is not surprising, and the fairly small number of “Other” responses ($n = 13[4.7\%]$) reassures the limited insecurity from the aforementioned problem of non-medical doctors replying to the survey. This is further confirmed by the primary workplace (see table 6.6), where the majority are employed at some hospital ($n = 162[58.1\%]$) or university ($n = 99[35.5\%]$).

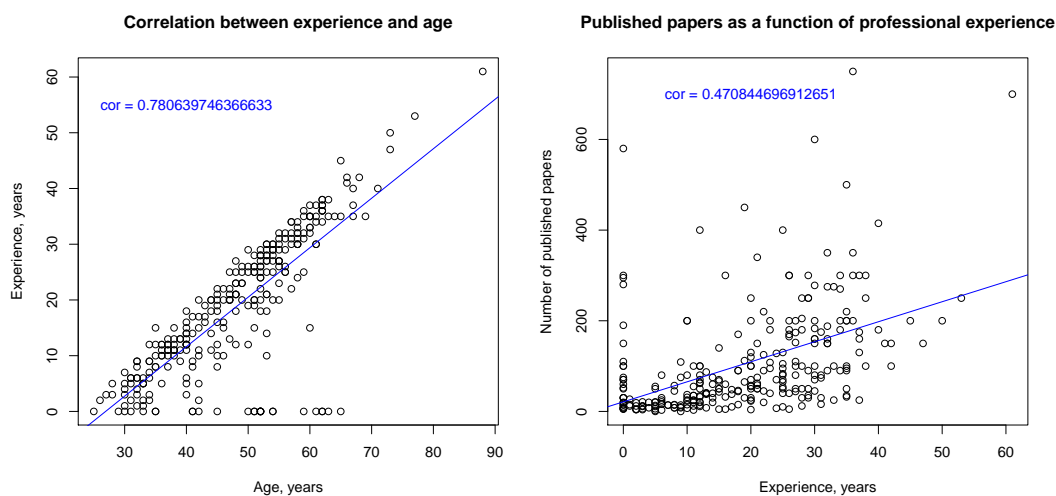
The distribution of specialties is displayed in table 6.7. Almost every specialty is represented, and a considerable amount are represented strongly; 21 specialties were selected by at least 5 participants, accounting for 203[72.8%] cases. It has not been pos-

²⁵According to statehealthfacts.org, based on a “Special data request on State Licensing Information from Redi-Data, Inc.”. <http://www.statehealthfacts.org/comparetable.jsp?ind=430&cat=8>

Table 6.4: Distribution of age and experience among participants. Experience is measured as years since completion of formal medical education. Both variables were recorded as exact years but are here cumulated in five-year intervals.

Age	<i>n</i>	%	Experience	<i>n</i>	%
20 – 24	6	2.151	0 – 4	43	15.41
25 – 29	33	11.83	5 – 9	30	10.75
30 – 34	30	10.75	10 – 14	43	15.41
35 – 39	36	12.9	15 – 19	24	8.602
40 – 44	36	12.9	20 – 24	34	12.19
45 – 49	58	20.79	25 – 29	42	15.05
50 – 54	37	13.26	30 – 34	31	11.11
55 – 59	28	10.04	35 – 39	22	7.885
60 – 64	10	3.584	40 – 44	5	1.792
65+	5	1.792	45+	5	1.792

Figure 6.4: Correlation between professional experience, measured as years since completion of formal medical education, and age in years. Linear regression is illustrated as a red line.



sible to identify an overall demographic for a suitable population, using the same categories. However, if the categories are compared to a report of the American Medical Association (Brooks, 2011), in which almost identical specialties are grouped into broader categories, the distribution of these categories correspond somewhat to the equivalent groups in the sample. The greatest difference is found in the group “Family medicine”, which is underrepresented in our sample, but also psychiatry, paediatrics, pathology and radiology are somewhat underrepresented, while the AMA group “Other specialties”²⁶ is somewhat overrepresented. It could be speculated that the underrepresentation of

²⁶containing Aerospace medicine, Dermatology, Emergency medicine, General preventive medicine, Neurology, Nuclear medicine, Occupational medicine, Physical rehabilitation and public health

Table 6.5: Primary job category

Job category	<i>n</i>	%
Research	154	55.2
Clinical practice/Physician/Surgeon	104	37.28
Administration	4	1.434
Management	4	1.434
Other	13	4.659

Table 6.6: Primary work place

Primary work place	<i>n</i>	%
General practice	3	1.075
Private company (e.g. Medicinal)	3	1.075
University	99	35.48
University hospital/Teaching hospital	133	47.67
Private hospital	9	3.226
Public hospital	20	7.168
Other public healthcare	10	3.584
Other private healthcare	2	0.7168

e.g. family medicine is associated with a high number of AMA members compared to a more modest research activity, as the specialties in the group (Family and General practice) are strongly focused on practice. The overrepresentation in “Other specialties” is likely to be explained by the very high number of participants from the “Public health” specialty. The relatively high number of participants selecting this specialty could be explained by the group of researchers with different background participating, as the post-graduate degree Master of Public Health is obtained by some researchers working in medical research with non-medical background. While we cannot be certain of this, the remaining sample is distributed roughly equally to the distribution of AMA-members and it appears reasonable to assume the sample to be a realistic representation of the contacted population, as it is likely the population would have similar differences from the AMA categories as the sample does. While there is some speculation in the above argumentation, which would not suffice for an exact representativity-analysis, the purpose is to identify if there are any extreme cases of bias in the sample. This does not seem to be the case.

The major part of the study participants came from the United States of America (see figure 6.5), however, this is connected to the amount of US-based e-mail addresses included in the survey population. Figure 6.6 shows the correlation between the number of participants from different countries and the number of e-mails sent to addresses with top-level domains²⁷ connected to that country. This analysis is sensitive to researchers using e-mail addresses different from their current country of residence, but also a little

²⁷e.g. .dk, .uk etc.

Table 6.7: Distribution of medical specialties among participants. Specialties represented by less than five participants were combined in the 'other' category

Medical specialty	<i>n</i>	%
Public health	21	7.527
Microbiology	16	5.735
Infectious diseases	15	5.376
Neurology	14	5.018
Pharmacology	12	4.301
Anaesthetics	11	3.943
Biochemistry	11	3.943
Cardiology	11	3.943
Paediatrics	11	3.943
Internal medicine	10	3.584
Gynaecology and obstetrics	9	3.226
Orthopaedics	9	3.226
General surgery	8	2.867
Oncology	7	2.509
Ophthalmology	7	2.509
Dental surgery	6	2.151
Endocrinology	5	1.792
Genetics	5	1.792
Psychiatry	5	1.792
Radiology	5	1.792
Urology	5	1.792
Other	76	27.24

unclear on specifically the US-based e-mail addresses, as the top-level domains .com, .edu and .gov are all connected to the USA, while other countries use country-specific codes or the international .org or .net. The latter are omitted in this analysis, while the number of e-mails sent to the USA are counted as the sum of all .com, .edu and .gov addresses, after the removal of generic domains as described above²⁸. The high correlation ($\rho = .964$) gives us an indication that there is a good connection between the contacted population and the drawn sample, with regard to country of residence, however, with a slightly higher emphasis on European countries.

In conclusion, the demographic data provides some evidence towards the representativity of the sample. The distribution of gender and country of residence in particular match the expected findings, while there is more doubt about the specialty and profession of the participants. Thus it would appear prudent to supplement the main factor analysis with a factor analysis of the demographic data as well, in order to discover potential confounding factors. Based on the above findings, we might expect the age,

²⁸i.e. gmail.com, yahoo.com and other international e-mail providers are omitted.

Figure 6.5: Number of participants from different countries. Only countries with more than two participants are displayed individually.

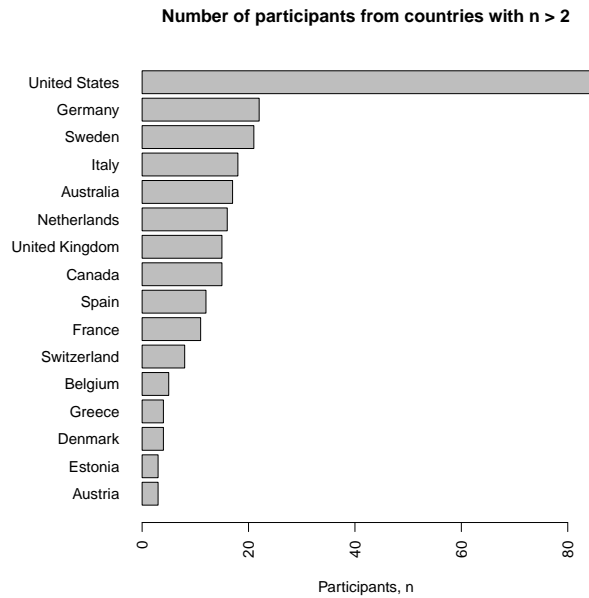
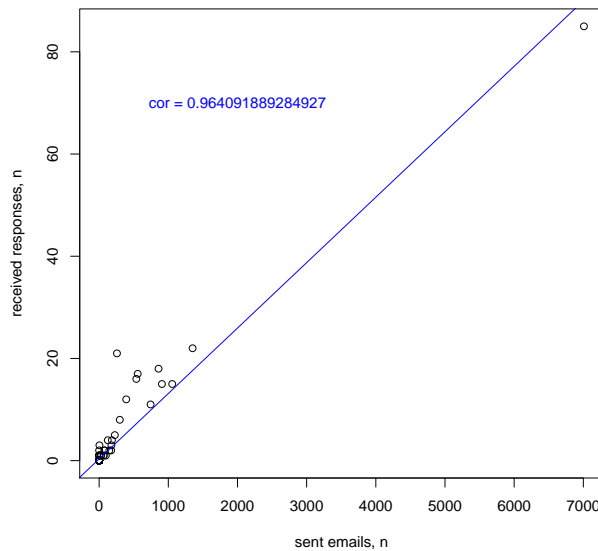


Figure 6.6: Correlation between number of participants per country and the corresponding number of e-mails sent to domains from those countries.



experience and scientific production of participants to be strongly correlated in a factor, while the remaining variables are not as closely connected and might not be sufficiently related to represent factors. As a result, the next chapter, containing the results of this

dissertation, will initiate the survey-section with an analysis of demographic factors. Before reaching this part of the analysis, however, the final step of conceptualising research quality will be described.

6.6. Creating narratives

In order to describe the dimensions of research quality, we need to qualify the identified variables and their common factors. For this purpose we used a circular approach, in which the identified factors, which are based on the survey variables and subsequently the interviews, were related back to the original interviews, thereby creating a qualitative narrative of each factors meaning (Fearon & Laitin, 2009). This means that the factor narration does not provide substantially new knowledge, as anything it can provide lies implicitly in the survey variables, however, the synergy effect of combining the qualitative and quantitative data may provide novel insights or combinations of insights that were otherwise obscured. The main goal of the narration was an improved interpretation of the results, which allows for more precise answers to the research questions.

The narration was performed manually, and thus there is an innate, subjective bias in the interpretation of the factors, however; the utmost effort was invested into representing all statements as equal, unbiased and objectively as possible. As the main purpose of the narratives is to serve for interpretation, this bias is not considered critical. Furthermore the variables sharing salient loadings on each factor will be included, in order to quantify the importance and agreement on statements which the variables were derived from.

The combination of interview statements, variable and factor scores, and narrated factors were finally compiled into a conceptual model of research quality, reflecting the identified dimensions.

7. Results

This chapter reports the results from the survey, first in their raw form, then followed by an analysis of the proposed data calibration procedures prior to the actual factor analysis. From this analysis the raw results, in shape of factor correlations, loadings and scores, are reported. These are further analysed in order to test the strength of the identified factors, and to discover relations to the other data from the survey. The narration of factors and the conceptualisation of research quality is reserved for the next chapter.

While the interviews are not included as results in this chapter - they were treated as midway results in the previous chapter - there is one particular statement which should be mentioned here, as it gives some justification for using published research as an evaluation-object; The most frequently occurring statement code, on average found almost twice in every interview, was that *there is a connection between the quality of published research articles and the quality of the research behind them*. In other words, if we evaluate the quality of research articles, we also gain an impression of the quality of the underlying research.

7.1. Data analysis

In this section, the data from the online survey, for use in the factor analysis, will be presented and analysed. The purpose is both to document the findings and describe them before further analysis, as well as decide whether the data is suitable for factor analysis. The initial part focuses on describing the raw data, which is followed by the results of raw score calibration and a test of the robustness of the calibration techniques used.

7.1.1. Raw survey data

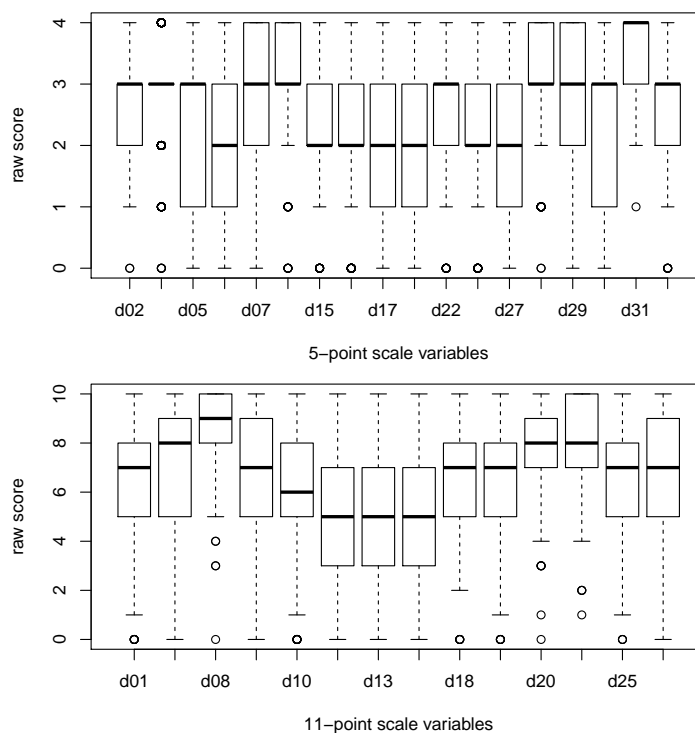
The survey data consists of three main sections, like the survey itself; demographic data as described in the previous chapter, the main variables and the article-specific variables. The main variables are measured on 5- and 11-point scales with assumed, simulated equidistance between categories. The assumption of equidistance may be too bold, to be exactly true, however; the scales were presented as radio-buttons with labels only at the highest, lowest and midvalues, thereby not letting semantic interpretation influence steps in values. The scales were created with minimum values at 0 (corresponding to “completely disagree”) and maximum values at 4 and 10 respectively (corresponding to “completely agree”). Both scales have an integer midpoint as well (2 and 5 respectively). The article-specific variables were not recorded as scores, as the variables were presented as topics for which participants could choose the three items they considered most important for the quality of an article, from a list of topically connected items. These variables are therefore recorded as frequencies, with the potential of weighting them according to the priority of selection.

To elaborate further on these two groups of variables, we will first plunge into the main variables. As these are the variables used for the factor analysis, they should be distributed in an appropriate way, which in this context means standard distributions. This is due to the common factor model operating with standard statistical procedures, such as means and variance. Also, it is very preferable if there is some variation between observations and variables, unless all variables are described by one common factor. In the following we will analyse this variance and the overall distributions’ likeness to the standard distribution. The latter will also be further analysed in the following section on calibration of the raw scores.

In figure 7.1 we can see the distributions of individual, raw variables. With very few exceptions, the entire spectrum is observable for each variable. Median scores range from the midpoint to the highest value. These distributions give an impression of inter- and intravariance, as we can observe very different median values between the variables, and we can observe very large distances between the medians and quartile ranges. The intravariance is illustrated more clearly in fig. 7.2, as distributions of variance for the two scale-types. As can be seen, the variance is spread from 0 to 2 for 5-point scales, with the majority in the 0.5-1.5 spectrum, and from 2 to 10 for the 11-point scales, with the majority in the 2-8 spectrum.

From table 7.1 we can see the variance of each variable, but we can also see the mean, median and standard deviation values. While the scales are ordinal and we should thus not use operators such as means, as values between ordinal classes are not represented in the ordinal space, our scales are using simulated equidistant classes. This means that we can interpret a mean value as the degree to which the average score lies between two classes. While this is not a perfect construction, and should not be performed from a very strict statistical viewpoint, there is a large body of literature showing the robustness of parametric operations on ordinal data, if the classes of values are well-made (Norman, 2010). We will therefore use these operations here, while carefully taking the ordinal

Figure 7.1: Boxplots showing the distribution of main variables measured on 5-point scales (top) and 11-point scales (bottom)



distributions into account. The means of the 5-point scale variables range from 2.00 to 3.63 and from 4.83 to 8.66 for the 11-point scale variables, showing also a great variance between the variables. While some of the variation between the variables can be ascribed to the different scales, it is clear that there is great variation between the variables. This observation is further supported when we plot the combined variable observations (still grouped by scale) as an empirical cumulative density distribution (see fig. 7.3). While not completely identical to a standard distribution, there is a clear tendency towards it.

Figure 7.2: Histogram of variance ranges for 5-point (left) and 11-point (right) scale variables.

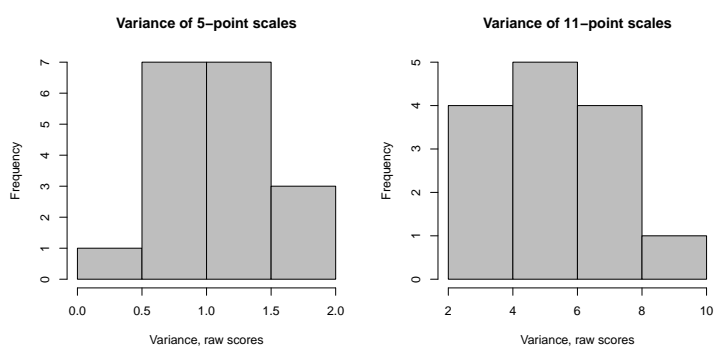
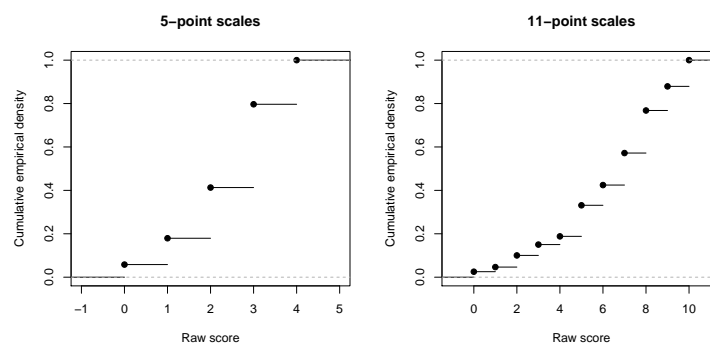


Table 7.1: Descriptive statistics for main variables. SD = standard deviation, var = variance.

variable	levels	mean	median	SD	var
authorqual	11	6.03	7	2.44	5.93
authorexpr	5	2.72	3	0.841	0.707
goodauthor	5	2.88	3	0.812	0.659
morejournals	11	6.76	8	2.8	7.84
limitproblem	5	2.28	3	1.22	1.48
limitgood	5	2.22	2	1.23	1.51
editorial	5	2.75	3	1.07	1.15
journalquality	11	8.66	9	1.45	2.12
datasource	11	6.73	7	2.53	6.38
jifqual	11	5.87	6	2.3	5.31
jifincent	5	3.09	3	0.903	0.816
selref	11	5.03	5	2.77	7.67
qualindep	11	5.32	5	2.68	7.18
guidequal	11	4.83	5	2.4	5.78
guideref	5	2.11	2	0.87	0.757
guideimplem	5	2.28	2	1.05	1.11
guidethink	5	2.12	2	1.04	1.08
citequal	11	6.4	7	1.85	3.42
citerespect	11	6.29	7	2.2	4.83
citetrend	11	8.02	8	1.62	2.63
citeref	5	2	2	1.06	1.13
citeping	5	2.57	3	0.994	0.987
citecoll	5	2.23	2	1.02	1.04
groundbreak	11	8.12	8	1.78	3.16
clinbas	11	6.82	7	2.24	5.02
baspurp	11	6.31	7	3.03	9.2
clinrel	5	2.09	2	1.4	1.96
gatekeeper	5	3.05	3	0.91	0.829
strictrev	5	2.7	3	1.14	1.3
innovatebias	5	2.41	3	1.24	1.53
negfind	5	3.63	4	0.584	0.341
goldstand	5	2.82	3	0.932	0.869

Figure 7.3: Empirical cumulative density distribution of 5-point (left) and 11-point (right) scale variables. All variables are combined into the two groups, creating a collective overview.



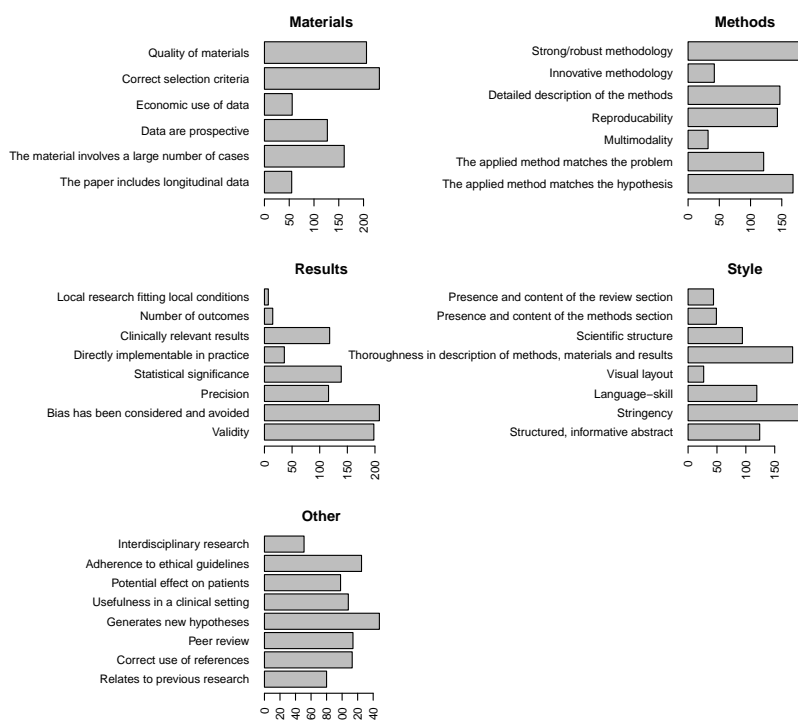
The main variables thus appear to be appropriately distributed for factor analysis; the variance between variables renders more than one common factor probable and the combined variables follow a standard-like distribution, allowing standard statistical procedures. This is illustrated both for the combined scores, but also for the individual variables (fig. 7.1). The distribution of the combined scores will be tested finally after the scores are calibrated. Before we reach this step, however; the article-specific variables are shown here in fig. 7.4. Each variable is plotted as an individual bar-plot, with values representing the frequency of the specific value being selected by survey participants. All frequencies are combined directly, whether the item was selected as first, second or third choice. A comparison was made with weighted counts, resulting in similarly ranked distributions, which is why we only show the pure counts here.

For the “other”-variable, the scores are somewhat evenly distributed, with two options scoring slightly higher than the remaining, while the other variables have more clear differences between scores. At this point, the frequencies are merely reported here, but they will be analysed in relation to the factor analysis in the following (section 7.2.4).

7.1.2. Calibration

As discussed in section 6.3, we need to calibrate our scales in order to be able to compare variables measured on 5 and 11 points, and we chose to compare the index and standard score calibration techniques. An important difference between the two calibration approaches (see page 88 for definitions) is the finite number of categories in index calibration as opposed to the scale values of the standard score calibration. In addition, the distance between categories in index calibration, when using multiple score ranges, is not static. This is not a problem when using index values in factor analysis, as any index values are still simulated equidistant - it is merely the categories which are not. When comparing index and standard calibration, however, this must be accounted for. To illustrate these differences, and also to test the suitability of the

Figure 7.4: Pure frequencies of response selections for article-specific variables.



common factor model for analysing the main variables, given their distribution, all calibrated scores were collected as one variable, and plotted as histograms (fig. 7.5) and empirical cumulative density distributions (fig. 7.6) respectively. In the histograms, the index categories have been combined into dodeciles, thus showing the distribution more clearly - albeit more coarsely - than the unevenly spaced categories would. A more fine grained distribution of the same values is shown in fig. 7.6, in the shape of the cumulative empirical density distribution of both calibration types. These figures give the impression, that the combined main variables resemble a Gaussian distribution, skewed toward the left.

The distribution is most clear for the standardised scores, however; as factor analysis operates on variance, we expect the calibrations to return similar factor loadings. If this is indeed the case, the index calibration should be used for further analysis, as it contains the most information - in fact it can be completely reversed to the original scores - but simultaneously, the standardised scores show us more clearly that the overall data is close to a Gaussian distribution; which is a requirement of the common factor model. Both calibration techniques intuitively appear to be sensible choices, but a robustness test in the shape of parallel factor analyses will reveal any potential differences. With regard to interpretation of the two calibration types, index values should be regarded as the degree of dis-/agreement for the variable, centered around 50, so that values closer to 0 show greater disagreement and values closer to 100 greater agreement. The value 50

Figure 7.5: Histograms showing frequencies of score ranges for index-calibration (left) and standard-calibration (right).

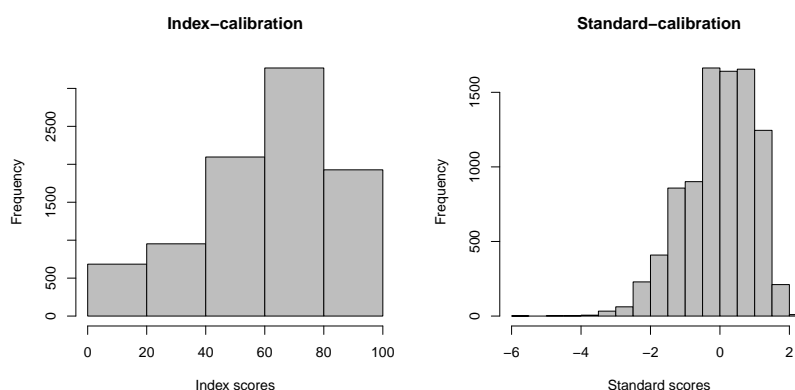
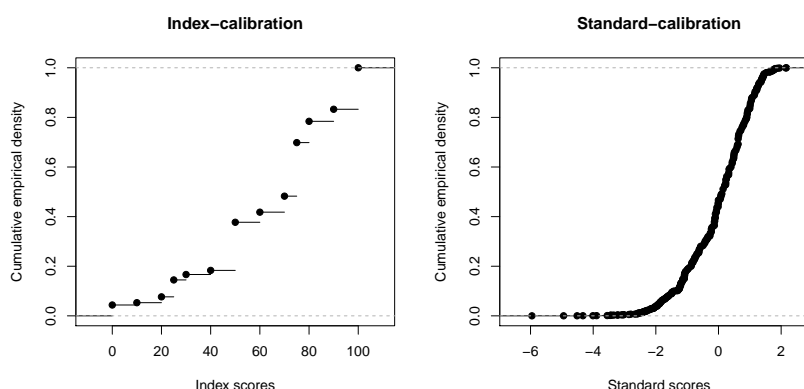


Figure 7.6: Empirical cumulative density distributions for all variables combined, using index-calibration (left) and standard-calibration (right).



itself indicates neutrality. The standard scores are measured as standard deviations from the mean, with values of 0 being exactly the mean, and ± 1 being 1 standard deviation above or below the mean.

7.1.3. Optimal coordinates for factor analysis

As stated previously (see section 6.4), using the common factor model requires input about the number of factors to find. In the following section, the results of the optimal coordinates approach for discovering the most probable number of factors (given the data) is reported for both calibrations, and an evaluation of the chosen factors, compared to other values, is performed.

The optimal coordinates approach uses Cattell's scree-test as a starting point. By comparing the superiority of experimental eigenvalues to predicted eigenvalues (determined through regression, see Raïche et al. (2013, p. 24)), one can determine the lowest

eigenvalue which is superior to the predicted, considered to be the optimal coordinate, n_{oc} . Eigenvalues, in factor analysis, inform us about how much variance each factor explains. Two variants for determining n_{oc} exist, using either Kaisers rule (7.1) or parallel analysis (7.2):

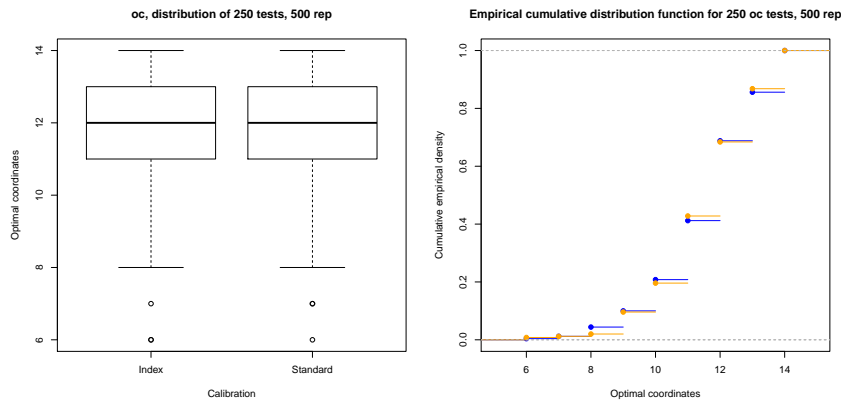
$$n_{oc} = \sum_i I[(\lambda_i \geq 1) \& (\lambda_i \geq \hat{\lambda}_i)] \quad (7.1)$$

$$n_{oc} = \sum_i I[(\lambda_i \geq LS_i) \& (\lambda_i \geq \hat{\lambda}_i)] \quad (7.2)$$

, where λ_i is the i th eigenvector, $\hat{\lambda}_i$ is the predicted i th eigenvector and LS_i is the so-called location statistic (see Raïche et al. (2013)). In this analysis, eq. 7.2 was used, with LS_i at the 0.05 quantile, as recommended by Raïche et al. (2013).

Parallel analysis requires several iterations, and different tests were performed to retrieve a stable n_{oc} . Using 50 tests with 50 repetitions each results in rather unstable results for both calibrations. Increasing the number of tests to 250 and the iterations to 500 produces far more stable results (see fig. 7.7), with median $n_{oc} = 12$ for both calibrations. As can be seen from the empirical cumulative density distribution, the variation between the two calibrations is very small, which strengthens our assumption of similar factor loadings for both calibrations.

Figure 7.7: Distribution of optimal coordinates (oc) with 250 tests, using 500 repetitions in each test, for both calibration types. To the left, boxplots show the overall distribution with quantile ranges, and to the right, the empirical cumulative density distributions for index-calibration (orange) and standard-calibration (blue) show the exact differences between the resulting optimal coordinates for the two calibrations.



When factor analysis was performed, using the common factor model and promax rotation, the resulting factor loadings for both calibrations were exactly identical. The calibration techniques can thus be considered robust, and per previous argumentation, we will use the index scores for the actual analysis.

7.2. Factor analysis

While the optimal coordinates approach gives a best guess as to the most probable number of factors, given the variance in the data, this guess needs to be tested, as does the choice of rotation. Therefore we will initiate this section with a confirmation of the factorial model, followed by the results from the actual factor analysis, a short description of the identified factors and a test of their significance through cluster analysis. The section is concluded with an analysis of the demographic and article-level variables seen in relation to the factors, in order to discover potential bias, confounders or underlying mechanisms. The narration of the factors from underlying interviews will be reported in the next chapter.

7.2.1. Model confirmation

When using promax variation, correlations between factors are allowed, however; if no, or only very low, correlations are present, we should use an orthogonal rotation instead, e.g. varimax. To test this, promax rotation was used, and the correlations between factors are listed in table 7.2. As can be seen, only very modest correlations can be found, and there is no ground for assuming substantial correlations to exist. Thus, the further analysis is performed using varimax.

Testing whether using 12 factors is in fact sufficient is equivalent to testing the χ^2 -statistic against the degrees of freedom, resulting in a p -value. In fig. 7.8 it can be seen how the use of 12 factors is the first case where the p -value exceeds 0.05, the standard level for statistical significance testing. The χ^2 -statistic is 190.13 on 178 degrees of freedom. Combined with the robust results from the optimal coordinates test, it would thus appear reasonable to maintain the 12 factors.

7.2.2. Factor loadings and descriptions

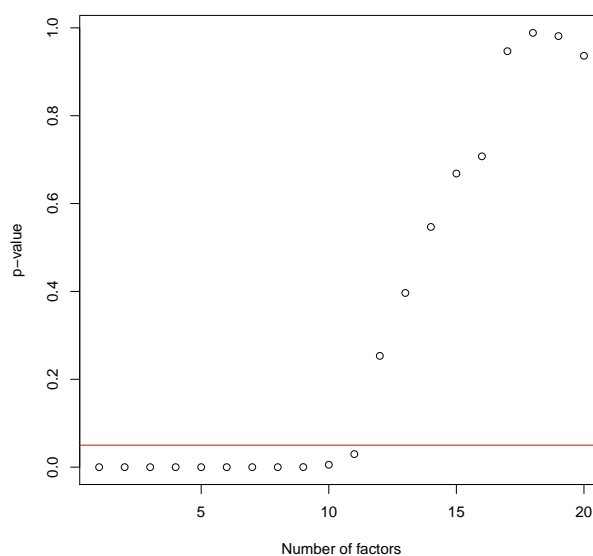
While the factors so far have been treated as equal, in terms of importance, it should be noted that they are not, nor are the variables. With regard to factor importance, one way of deciding which factors contribute most to explaining the data is by looking at the proportion of total variance explained by the factor. These values are displayed in table 7.3, showing that the overall variance explained by the factors is only moderate (with a total of 46.9% explained by the 12 factors), but that the first three factors contribute a considerably higher amount than the rest, and the final two contribute a very small proportion. In the further analysis we should therefore emphasise the factor importance accordingly. In the next chapter, where factors are narrated, all factors will be included, as will all variables, but emphasis will be put on those considered more important; for factors depending on the variance they explain, and for variables depending on their distribution.

The final factor loadings are displayed in table 7.4, omitting those loadings between -0.20 and $.20$. The order of variables have been rearranged to provide a more clear

Table 7.2: Correlations between factors. Factor numbers are used as labels, as the correlation matrix is symmetrical, only the upper half is displayed.

Factors	1	2	3	4	5	6	7	8	9	10	11	12
1	1											
2	-0.18	1										
3	-0.071	-0.16	1									
4	-0.12	-0.13	-0.13	1								
5	-0.17	-0.13	-0.014	-0.11	1							
6	0.041	0.0049	0.061	0.026	-0.018	1						
7	-0.063	-0.078	-0.06	0.014	-0.08	-0.017	1					
8	-0.12	-0.076	-0.055	-0.15	-0.11	-0.041	-0.046	1				
9	7e-04	-0.11	0.072	-0.073	-0.13	-0.1	-0.028	-0.12	1			
10	-0.05	-0.049	-0.03	-0.063	-0.061	0.015	-0.095	-0.17	-0.079	1		
11	0.079	-0.28	-0.26	-0.05	0.039	0.042	0.013	-0.042	0.02	-0.098	1	
12	-0.2	-0.14	-0.011	-0.062	-0.21	0.0063	-0.051	0.041	-0.056	0.066	0.12	1

Figure 7.8: p -value for testing the number of factors, as a function of the number of factors, using Bartlett's test and varimax rotation. The red line marks the 0.05 boundary.



	Factors											
	1	2	3	4	5	6	7	8	9	10	11	12
Pvar	.058	.056	.052	.039	.039	.039	.039	.036	.035	.033	.023	.020

Table 7.3: Proportion of total variance (Pvar) explained by individual factors.

association between variables and factor loadings. For the construction of salient factors, an absolute boundary of .30 was chosen, resulting in common factors for almost all variables, and fairly strong loadings in each factor. There are no clear rules for what is considered a good salient boundary, and it is here chosen based on the distribution of loadings. The choice of a higher value would have resulted in much fewer and less diverse factors, while a lower value would have created too much overlap between factors, with very low internal commonality.

Conceptualising research quality in medicine for evaluative bibliometrics

Variable	Factors											
	1	2	3	4	5	6	7	8	9	10	11	12
qualindep	-0.33											
journalquality	0.44											
jifqual	0.76											
jifincent	0.62										0.2	
selref	0.45		0.36								0.26	
citeref	0.26		0.54									
citeping			0.72									
citecoll			0.66									
guidequal		0.78										
guideref		0.65										
guideimplem		0.53										0.2
limitproblem				0.89								
limitgood				-0.57								
gatekeeper					0.53						-0.28	
strictrev					0.68							
clinrel		0.22			0.41	0.32						
baspurp						0.97						
goldstand					0.39							0.31
negfind												0.37
clinbas									0.2			0.31
authorqual							0.67					
authorex	0.22						0.5		0.21			
goodauthor							0.54				0.24	
citerespect			0.24					0.3	0.21			
citetrend								0.98				
citequal	0.21	0.23							0.92			
innovatebias					-0.21					0.95		
morejournals											0.49	
guidethink		-0.4									0.31	
editorial	0.23					0.27						
datasource												
groundbreak			0.23									0.23

Table 7.4: Factor loadings for main variables. Only loadings $\geq .20$ or ≤ -0.20 are displayed for ease of reading. Salient loadings are marked with bold text.

To evaluate the isolated factors, we can scrutinise the meaning of the variables in the shape of the questions which were presented to the participants. Doing so, it should be clear whether the identified factors carry meaning, or are random combinations of variables. In the list below, each factor has been assigned a label and a description, based on the content and meaning of the associated questions:

Factor 1 - Journal prestige Variables relating to the perceived quality and prestige of journals, as well as the effect from journal impact factors.

Factor 2 - Clinical guidelines Variables relating to the use and meaning of clinical practice guidelines.

Factor 3 - Referencing behaviour Authors own behaviour with regard to referencing.

Factor 4 - Method section The length of method-sections.

Factor 5 - Subjective quality Variables relating to the subjective quality of research, such as peer review and clinical relevance.

Factor 6 - Basic to applied Variables relating to the purpose of basic research, and the relevance of clinical research.

Factor 7 - Author Variables relating to the authors of research articles.

Factor 8 - Citation meaning Variables relating to the meaning of citations.

Factor 9 - Citation quality Variable related to the connection between citations and research quality.

Factor 10 - Innovation stunt Peer review stunting innovative and groundbreaking research.

Factor 11 - Skepticism Variables related to an overflow of journals and skepticism towards clinical practice guidelines.

Factor 12 - Propriety Publishing of negative findings and the use of certain golden standard methods

Most of the factors make obvious sense, and the combination of variables correspond well to what one might expect, e.g. for authors and journals. It might be surprising and very interesting to see that the citation-related variables are spread across several factors, in particular two factors dealing with the participants' own referencing behaviour, and the assumed meaning of citations. Citations, and their link to quality, appear as a single-variable factor, and also factor 10 is only connected to one variable. While these factors could be omitted, we will maintain them for further analysis, in the following section as well as the next chapter, allowing potential insights to be revealed subsequently.

7.2.3. Factor evaluation

Before the demographic analysis and narration of the factors, it is prudent to further evaluate the factors through cluster analysis, thereby also allowing us to identify groups of participants for the demographic and article-level analysis. If the factors are sufficiently salient, it should be possible to identify clusters of observations with unique distributions of responses. For this purpose we use the Bartlett scores, extracted from the factor analysis, representing the factors' influence on each observation. The Bartlett scores correspond in interpretation to the previously used standard score (0 = mean, 1 = one standard deviation). To cluster the observations, vectors for each observation, containing the Bartlett scores on the 12 factors, are compared for similarity. Several procedures exist for this purpose, but as our aim is to compare the scores on each factor, using scale data, the cosine similarity measure (eq. 7.3) provides a useful and robust result (Schneider & Borlund, 2007).

$$\text{cosim}(A, B) = \frac{\sum_i a_i * b_i}{\sqrt{\sum_i a_i^2 \sum_i b_i^2}} \quad (7.3)$$

These vector comparisons can subsequently be ordered in a similarity matrix, from which clusters can be extracted. Again, several approaches exist, and as this is not the core purpose of this dissertation, we will be using the straightforward, but also robust, approach of hierarchical, agglomerative clustering. The result of this procedure can be plotted as a dendrogram, from which a suitable level for cluster construction can be selected. This level should result in a useful number of clusters, but should also be clearly discernible from other solutions, i.e. it should not be an arbitrary choice between n , $n + 1$ or $n - 1$ clusters; our n needs to be clearly discernible from the other solutions. Such a solution is displayed in fig. 7.9, with the minor variation that the cosine similarity has been transformed to cosine dissimilarity ($1 - \text{cosim}$) in order to let R plot the dendrogram. Clusters are shown as red boxes.

We further evaluate the found clusters by plotting their mean Bartlett scores for all factors in a heatmap (fig. 7.10). From this we can see clear differences between the clusters, especially from the top- and bottom-scores. Combining these scores with the factor labels, we can describe the clusters as in table 7.5, using -0.5 and $+0.5$ as thresholds for low and high scores respectively. We do see that there are some clusters of observations, but there does not appear to be an obvious system therein. Nonetheless, the cluster analysis confirms that the factors can produce groups of observations with considerably different distributions of scores. To further test whether there is an underlying structure behind the clusters, they were used in the following analysis of article-level and demographic variables.

7.2.4. Article-level and demographic analysis

An analysis of the identified factors seen in relation to the demographic variables and the article-level variables respectively was performed. The purpose of the analysis

Figure 7.9: Dendrogram showing the similarity of observations, as likeness of Bartlett scores. Similarity is shown as distance, i.e. values close to 0 are very similar. Clusters are shown as red outlines.

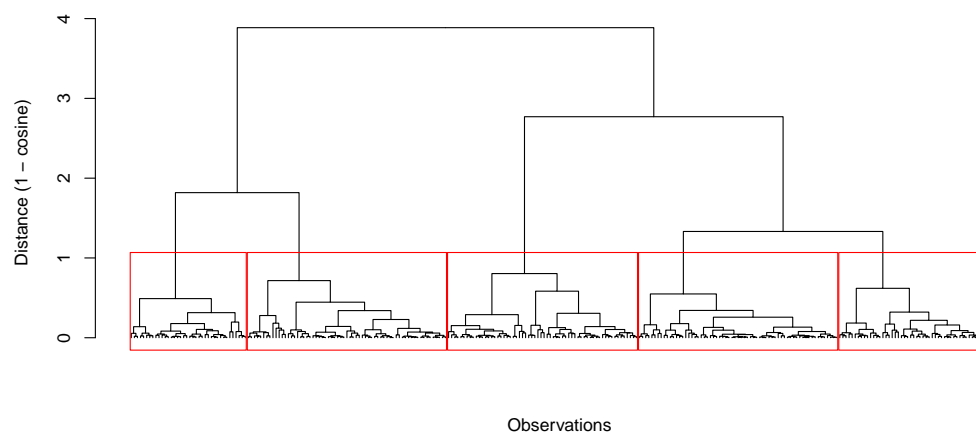


Figure 7.10: Heatmap showing the average Bartlett scores for each cluster on each factor.

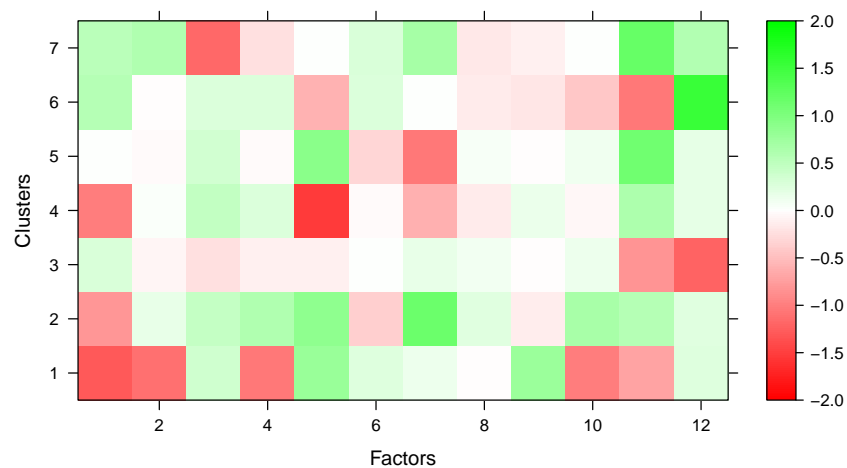


Table 7.5: Low- and high-scoring factors in each cluster, using thresholds of ± 0.50 .

Cluster	Low scores	High scores
1	Journal prestige; Clinical guidelines; Method section; Innovation stunt; Skepticism	Subjective quality; Citation quality
2	Journal prestige	Method section; Subjective quality; Author; Innovation stunt; Skepticism
3	Skepticism; Propriety	
4	Subjective quality; Author	Skepticism
5	Author	Subjective quality; Skepticism
6	Subjective quality; Skepticism	Journal prestige; Propriety
7	Referencing behaviour	Journal prestige; Clinical guidelines; Author; Skepticism; Propriety

was to uncover whether any of these variables correlated, indicating connections between the factors and these variables. This was done using the Bartlett scores, similarly to the cluster analysis above, however, no strong correlations were found for any of the variables. A comparison between clusters of observations and the article-level and demographic variables did not reveal any strong correlations either, and as only weak correlations, or moderate correlations based on very few observations, were present, it did not seem feasible to pursue this analysis further.

7.2.5. Summary

Over the course of this chapter, the results from the main and article-level variables have been reported and analysed. The main variables were found to be appropriate for factor analysis, which was subsequently performed using varimax rotation (as no significant correlation was identified between factors). 12 factors were identified, and based upon a cluster analysis of the original observations and the factor scores, the factors were also found to bear considerably different meanings. A brief analysis of the content of the variables sharing common factors also showed that the factors were intuitively sensibly constructed. The analysis of demographic and article-level variables grouped by either factor scores or clusters did not reveal any strong correlations between these variables and the factors. While this is the case, the article-level variables still provide information on article-specific attributes believed to be related to research quality. The next chapter focuses on narrating the identified factors, based on their connection to the original interview statements, codes and full interviews, as well as constructing the conceptual model of research quality intended to summarise the findings of this dissertation. While the focus is on the factors, the article-level variables will be used to elaborate and specify in publishing-related areas.

8. Factor narration and conceptualisation

As described in the methods chapter 6, the purpose of narrating factors is to tie the quantitative data to qualitative descriptions, thereby qualifying the findings and improving the interpretation of identified factors. This process allows for a more elaborate conceptual model to be created subsequently. The exact narration process was explained in section 6.6. In some cases, quotations from the interviews are listed in the narratives. These quotations are the authors best attempt at providing a direct translation from Danish to English, with only minor rephrasing to transcribe from spoken to written word. The narrations are followed by the conceptual model. The relationship between the research quality concept and bibliometric measurement is also discussed in this chapter, and further discussion of the results and their implications for evaluative bibliometrics follows in the next chapter.

From the interviews we gained an impression of how research quality was perceived in medical research, the first research question of this dissertation, and through the narration of the quantified factors we aim to provide the final evidence for an answer to this question. Through the construction of a conceptual model the second question, concerning the relationship of research quality dimensions, is sought to be answered. With regard to the final research question, it will be treated following the final response to the other two.

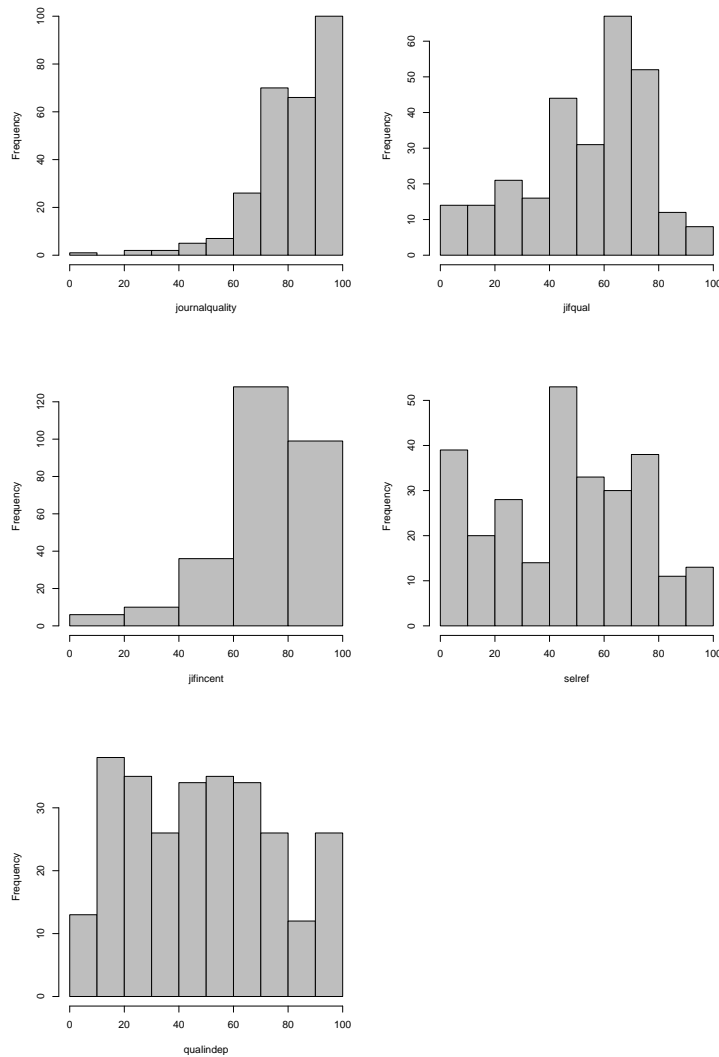
8.1. Narrated factors

Throughout these factor narrations, the index values of variables with salient factor loadings for the respective factor are reported as histograms. When variables are compared to the narratives, v is used to indicate the values of said variable, \bar{v} the sample mean of variable v , $n(v \geq x)$ the number of observations larger than x and $P(v \geq x)$ the proportion of observations larger than x from the entire sample of observations.

8.1.1. Factor 1 - Journal prestige

Index-scores for this factor are shown in figure 8.1.

Figure 8.1: Histograms of index scores for variables associated with factor 1 - Journal prestige



Several different opinions about journal quality, prestige and esteem were collected, a number of which are also connected to peer reviewing and citations. The most dominant topic related to journals was the use and meaning of the journal impact factor. Some of the interviewees demonstrated detailed insight into some of the problems of the journal impact factor, such as field differences, and the missing connection between journal impact and expected article impact. Some speculated that the journal impact factor reflects utility rather than quality, however; the majority believed the journal impact factor to be indicative of the quality of the journal and as a result the articles published therein. On the other hand, different mechanisms behind high journal impact factors, and the type of research published in the highest impact journals were proposed as well; non-scientific criteria, such as media attention or news value are used by editors of high-impact journals, but also clinical research is supposedly easier to publish in high-impact journals, than basic research. Fraud and scientific misconduct was also mentioned as more likely in high-impact journals, although the interviewees also believed the peer reviewing of high-impact journals to be better. Almost all interviewees who referred to specific examples of high-impact journals referred to general medical journals and not specialty journals. In particular *The Lancet* and the *New England Journal of Medicine* were used, but also the multidisciplinary journal *Nature* was mentioned. One interviewee stated:

“It would be great to be published in *The Lancet* - everybody wants to be published in *The Lancet*.” (Interview M19)

While several of the interviewees were skeptical towards the meaning, use and over-use of journal impact factors, they were all aware of them and using them. Other interviewees reflected similarly on both the use of journal impact factor as a “language”, and its necessity when applying for grants, but also for general career advancement. One interviewee said:

”[Journal Impact Factor] is a means of communication in my world, whenever I need funding for my research.” (Interview M04)

This was especially seen as negative by researchers in areas where journals did not reach high (raw) journal impact factors, as they felt disadvantaged when applying for grants, regardless of the qualities of their work. This was also the case for those working with geographically dependent diseases²⁹, as they found it much more difficult to publish e.g. Scandinavian relevant results in American journals - which generally have higher journal impact factors.

With regard to low-impact journals, interviewees believed that articles therein were more diverse, both in regard to quality and type. As an example, it was said that articles in low-impact journals may be of good quality, but are less spectacular than those in high-impact journals. This was also given as a reason for why some journals are more prestigious than other; articles therein provide more definite answers or results, and

²⁹e.g. viral Hepatitis, which occurs as five different types

may change the way we think about things. Also, the low-impact journals were considered more likely to publish low quality articles. On the other hand, some interviewees believed that truly groundbreaking research might sometimes be more likely to be published in low-impact journals, as it might be too much outside the mainstream, or too unexpected, to be published by high-impact journals, as it might lead to suspicions of fraud or flawed results. Anecdotal evidence was given to support this view, although not specific enough to be documented here. The concept might be akin to “sleeping beauties” in citation analysis (van Raan, 2004), and is opposed to the previously mentioned view of articles in prestigious journals changing the way we think. Both effects can co-exist, however.

Several interviewees mentioned the self-feeding mechanism of journal impact factor: more articles are submitted to high-impact journals, as they are more desirable to publish in, and these journals have a larger base to pick high-quality articles from, which they expect to have an impact and be cited. They also stated, that respected, famous or popular authors were more likely to be published in high-impact journals, citing both cronyism and author skills as mechanisms. As a final factor in the feeding mechanisms for journal impact factors, they mentioned that they would be more likely to select references from high-impact journals in their own research.

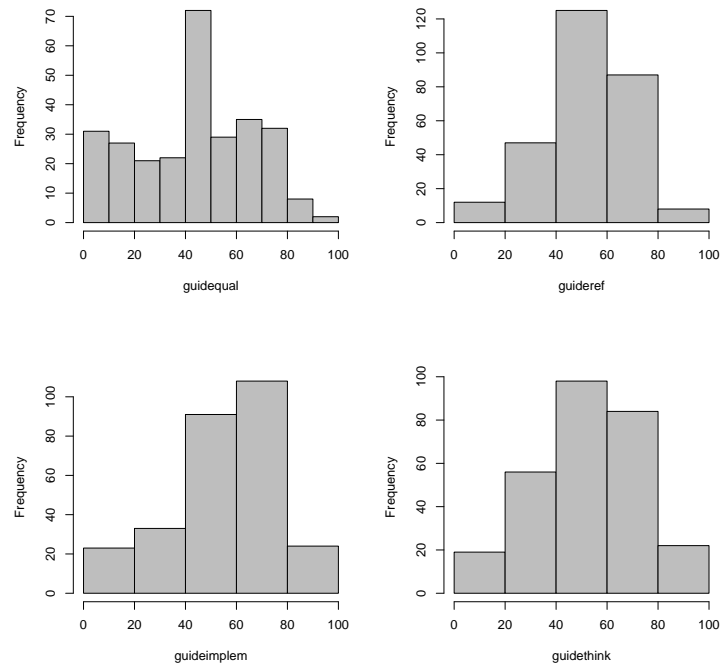
With regard to the survey variables, there was a high degree of agreement on the `journalquality` and `jifincent` variables, meaning that almost all participants agreed strongly ($\bar{v} = 86.6$ and $\bar{v} = 77.2$ respectively) on 1) some journals publish articles of higher quality than others and 2) high impact factors are an incentive when selecting a journal to publish a paper in. The journal impact factor was also believed to be indicative of the quality of articles in a journal, although not as strongly ($\bar{v} = 58.7$). As for using the journal impact factor when selecting references, the replies were very diverse ($n(v < 50) = 101$, $n(v = 50) = 53$, $n(v > 50) = 125$, $\bar{v} = 50.3$). Views on article quality being independent from journals are negatively correlated with the other variables in this factor ($\rho = -0.33$). In conclusion, the survey confirmed the views presented in the interviews, that journals are indicative of article quality; that journal impact factors are important in medical research; and that the journal impact factor is not a perfect measurement.

8.1.2. Factor 2 - Clinical guidelines

Index-scores for this factor are shown in figure 8.2.

Not many interviewees mentioned clinical guidelines, and very different opinions were presented. During the pilot interviews, a feeding mechanism from basic to clinical research and to clinical guidelines was mentioned by one participant, and this thought was presented to some participants in the main interviews. Those who were presented with the idea reacted affirmatively, although some suggested it was too simple to represent reality. One participant described clinical guidelines with disdain, claiming they are based on “what we think we know” rather than real evidence, while other participants believed the references selected for clinical guidelines reflected the highest evidence on

Figure 8.2: Histograms of index scores for variables associated with factor 2 - Clinical guidelines



the topic. Participants proposed there might be specialty-differences for the quality of clinical guidelines.

Some regarded the inclusion of a study in a clinical guidelines as a proof of quality, and clinical guidelines were by some seen as the highest level of evidence available. Several participants stressed the need for high evidence studies, systematic reviews and meta-analyses for clinical practice.

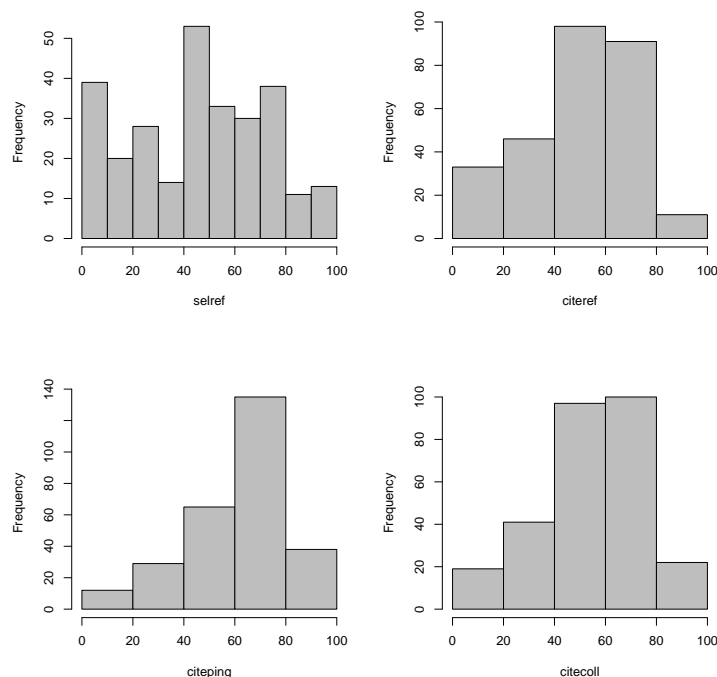
In the survey, very diverse scores were collected for all three variables in the factor, with an overweight around the midvalue. Mean values were $\bar{v} = 48.3$ for *guidequal*, $\bar{v} = 52.9$ for *guideref* and $\bar{v} = 56.9$ for *guideimplem*. This indicates that participants on average agreed slightly more on the importance of guidelines for implementing research in practice, but were rather indifferent with regard to clinical guidelines as indicators of research quality. The *guidethink* variable, related to the statement on guidelines being based on what we think we know, is negatively correlated to the other variables in the factor.

8.1.3. Factor 3 - Referencing behaviour

Index-scores for this factor are shown in figure 8.3.

Several interviewees mentioned effects akin to the Matthew-effect, e.g. by saying “once you have many citations, people will keep citing you.”, and in the same line, doing groundbreaking research was mentioned as a reason for achieving the first, many

Figure 8.3: Histograms of index scores for variables associated with factor 3 - Referencing behaviour



citations. As a mechanism behind the Matthew-effect, several participants mentioned using citations as an initial relevance-control; if a paper was cited previously, they were more inclined to read and use it. Also in the further process, citations play a role, as participants said they would be more certain that using a previously cited reference would be accepted by peers, and one mentioned that as a peer reviewer, they would notice unexpected references. Both normative and rhetorical mechanisms may be behind this type of behaviour - the use of specific references for specific purposes, e.g. methods, regarded as the norm in a research field, will fit into the Mertonian paradigm for referencing theory, which is something a peer reviewer might notice. On the other hand, the use of references solely on the grounds of the number of citations received is more a question of rhetorical use, or “persuasion” rather than the Mertonian “reward”.

As for using previously cited references, this was perceived as a sort of quality control by some; if they used the same references that others did, they could be fairly certain they had found the relevant literature. Participants related this to an overflow of information:

“I would be more inclined to look at those articles others cited... ..rather than look at a hundred articles and hand-picking the best” (Interview M15)

. There was a distinction between using reference lists as check-lists and reusing reference lists, however:

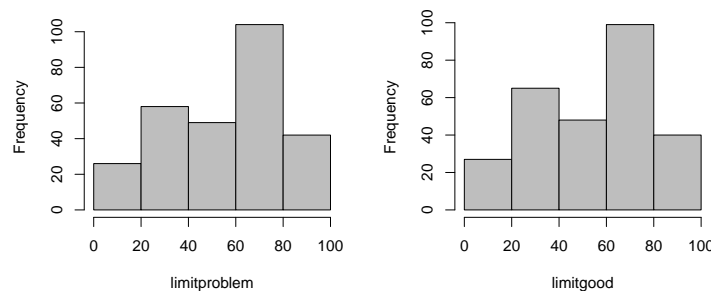
“it may be they choose [an article] which has been cited before, and they think it’s a good article to cite. So they cite it too, and just read the reference lists of other articles, or those articles which use the same reference. Then they don’t need to order and read it - that is bad quality.” (Interview M03)

The variables included in the factor moderately reflect these statements. The `self-ref` variable, shared with factor 1, is inconclusive and mostly related to the journal impact factor. With regard to selecting previously cited articles, articles by well-known researchers and the work of colleagues, survey participants were most inclined to cite well-known authors ($\bar{v} = 64.2$), followed by citing colleagues ($\bar{v} = 55.8$) and finally previously cited work ($\bar{v} = 50.1$). While the selection of references signals something about the quality of the cited works, the type of quality depends on the general reason for citing. It would appear that persuasion does play a role, but not a decisive one. The reasons for citing well-known authors - which scores highest - might signal that some authors generally produce higher quality work, which is recognised and rewarded by the research community. There is no clear evidence of this either, but since the citation counts of references was seen as less of an incentive for selecting references, it is not likely that authors are selected more purely on the basis of citation counts. It is thus not possible to conclude whether citation counts represent a persuasion or reward effect from these data, but it is likely to be a mix of the two, and it is therefore likely that citations represent a communicative quality as well as a quality of the underlying research.

8.1.4. Factor 4 - Method section

Index-scores for this factor are shown in figure 8.4.

Figure 8.4: Histograms of index scores for variables associated with factor 4 - Method section



The most commonly referred issue about the method section was the need to write reproducible articles. However, with regard to limiting the length of articles, and this having an effect on the length of the method section, two opposing statements were made by interviewees. One stated that having good figures was far more important than the method section, as “no-one reads the method section anyway”. Opposed to

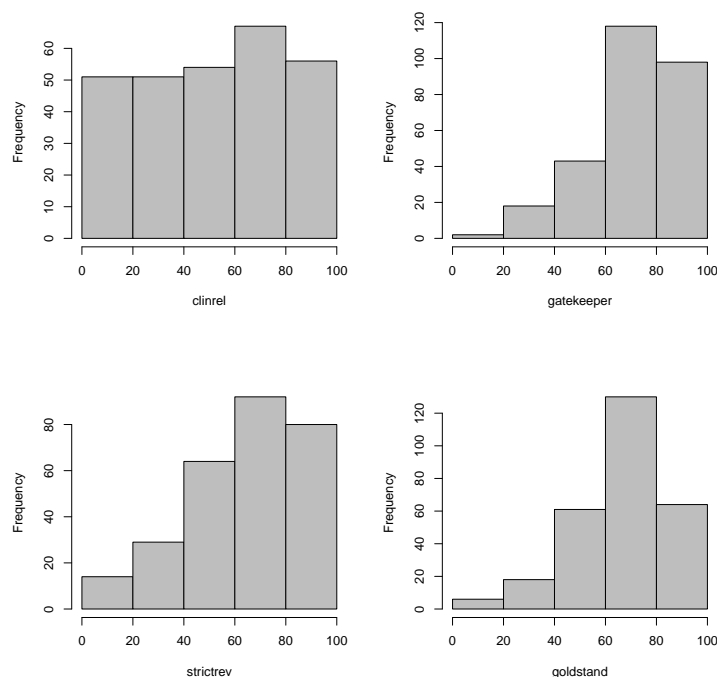
this was a critique of some journals, which limit the length of method sections, and move this section to the end of the article. The interviewee stressed the importance of having enough space to document procedures for others to reproduce it and had found several cases in which this was not possible because of short method descriptions.

This factor consists of two negatively correlated variables, `limitgood` and `limitproblem`, both representing the two opinions stated above and illustrating the reciprocal relationship between them.

8.1.5. Factor 5 - Subjective quality

Index-scores for this factor are shown in figure 8.5.

Figure 8.5: Histograms of index scores for variables associated with factor 5 - Subjective quality



The factor focuses on peer review, clinical relevance and golden standards for research methods and study designs, which can together be seen as a type of subjective quality. It could also be labeled as a content quality, as it focuses on the quality of the content in the article.

The topic of peer review sparked very different reactions in the interviewees, but overall they agreed that some sort of peer review is necessary as a quality insurance tool. Several flaws were mentioned, including the lack of time for peer reviewing, resulting in sub-standard reviews; the sometimes editorial role of peer-review, where articles are

discarded because of lack of space³⁰ in print journals; and too varying quality of peer review, e.g. seen by strongly diverging reviews of the same manuscript. Many participants considered peer review to be more strict in prestigious journals, and one participant wondered about where the plethora of new, small open-access journals would find reviewers, as most researchers are already busy peer-reviewing. Procedures such as paying peer reviewers or using open peer review procedures³¹ were suggested as means of improving the traditional peer review system. With regard to research quality, peer review was suggested to be a gatekeeper, preventing the lowest quality articles from being published, however; it was not seen as an actual quality indicator. Also in the survey, there was a clear tendency to trust in peer review as a gatekeeper ($P(v > 50) = .77$) and there was also overall agreement on more strict peer review leading to higher quality published research ($P(v > 50) = .62$).

Clinical relevance is used as a collective term, for slightly different wordings by interviewees, who otherwise agreed on the meaning - the usefulness of research in a clinical setting, implications for patients and direct translatability into practice. Clinical relevance was perceived far more important by interviewees than by the survey participants, however; the survey only reflected a specialised part of clinical relevance. The variable `clinrel` stated that all research - whether basic or clinical - should be clinically relevant. The variable scores are distributed over the entire range of scores, as can be seen in fig. 8.5, and different opinions on specifically this phrasing of the topic can be found in the interviews as well. While some participants argued that for basic research to have any value, it should at some point be relevant to the clinic; others argued that basic research should be free of such constraints as the real changes take place in basic research areas. In a more broad sense, however; clinical relevance was seen as incredibly important. When asked if there was a thing as irrelevant knowledge, one interviewee replied:

Yes, in my field there is. I am a doctor, so when seeing myself as a part of a system, with the purpose of keeping people as healthy as possible and curing sick people, then there is definitely something that is more relevant than other. So I might be biased by my background. When we do research in my field, we tend to think, "What can it be used for?", even when we go into basic research, it's still there; Could you imagine at some point, this could help some people?" (Interview M04)

Some interviewees went further into detail, stating that effect sizes and the number of patients a procedure could help was a quality indicator for medical research, and that the effect on health was the real impact of research.

This clinical relevance is clearly related to the golden standards, or the evidence-system, in which randomised controlled trials are often seen as golden standards for

³⁰While this was considered peer review by the interviewee, it is more likely that it is in fact a question of editorial work

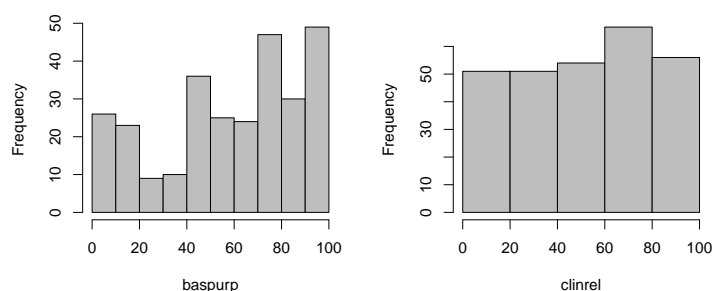
³¹e.g. publishing review notes, or allowing public comments.

performing treatment-related research. However, several interviewees pointed out the importance of other study designs, and the need for different standards in different settings. Several interviewees argued that the most important quality with regard to study design was a good connection between the research question and the study design, indicating this is not always the case. On the most general level, there was a distinction between basic and clinical research evidence. Where the number of patients involved in a study is crucial for many clinical studies, basic research might as well use few samples and multimodality to collect strong evidence. Regarding clinical research, golden standards for research methodology were seen as a help to clinical practitioners, both in the medical and surgical fields. The need for golden standards appeared different for researchers, as e.g. case studies, which are usually seen as low evidence study designs, are very important for generating new hypotheses, and thus potentially much more interesting for researchers. One interviewee stated that their definition of what makes an article good would change whether they were doing rounds or research. Summing up, golden standards for methodology were generally seen as important, but a flexible view upon what they are golden standards for was stressed. In the survey, there was an overall agreement on the preference of some methods as golden standards ($P(v > 50) = .70, \bar{v} = 70.4$).

8.1.6. Factor 6 - Basic to applied

Index-scores for this factor are shown in figure 8.6.

Figure 8.6: Histograms of index scores for variables associated with factor 6 - Basic to applied



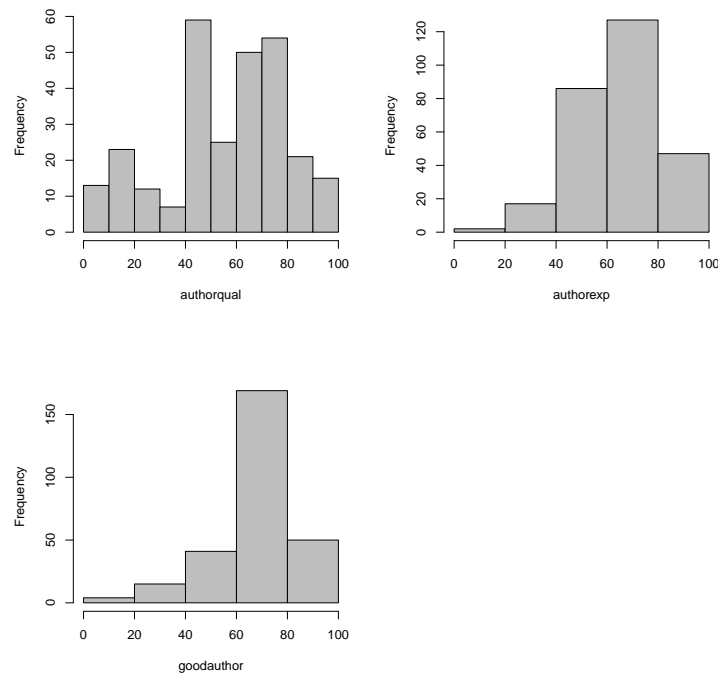
The `clinrel` variable from factor 5 is also present in this factor. The other variable, `baspurp`, is also related to some of the viewpoints presented in the previous factor description, as it is formulated as: “The ultimate purpose of basic research is improved health”. This variable adds some nuances to the `clinrel` variable, as it uses “improved health” rather than “clinical relevance” as the final endpoint - two terms which are strongly related but not identical. Interestingly, there is a somewhat stronger agreement on this variable in the survey ($P(v > 50) = .63, \bar{v} = 63.1$), but also a rather strong opposition ($P(v \leq 20) = .18$). As most interview statements concerned

with the above were covered in factor 5, we will only expand it slightly here, including an observation which combines precisely these two variables. The aforementioned feeding mechanism from basic to clinical research was by one participant elaborated as a process in which discovery goes from observations in e.g. case studies, generating hypotheses, which are then related to basic research findings, resulting in theoretically founded hypotheses, which can in turn be tested in clinical trials. In this model, basic research exists in its own right, parallel to clinical research. While there appears to be broad agreement on basic research being important and a prerequisite for much clinical research, it is also seen as largely irrelevant to clinical practice. All in all, there seems to be an agreement, that basic research should exist in its own right, that it is important for the progression of clinical research, but also that while the final goal may be improved health; basic research may be so remote from practice that the connection does not seem apparent or obvious.

8.1.7. Factor 7 - Author

Index-scores for this factor are shown in figure 8.7.

Figure 8.7: Histograms of index scores for variables associated with factor 7 - Author



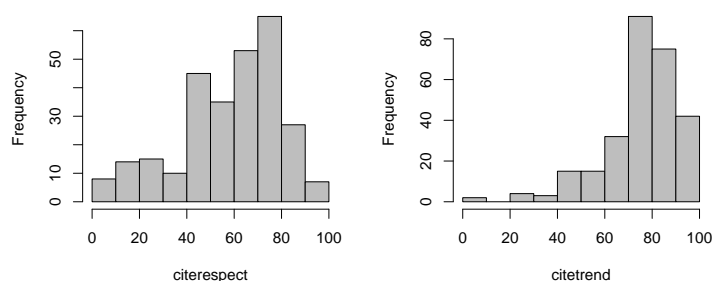
Several interviewees mention authors, during the presentation of the papers they brought for the interview, which they regard quite highly. From their statements it appears that they generally trust these authors to provide good or even excellent quality research. Some interviewees also directly state that some authors produce better research

than other. As an example, it is pointed out that some authors write articles to bolster their resumé, while other do it because they have a message; and the latter is clearly preferred by those participants issuing the statement - the former is regarded as low quality research (and might be categorised as scientific misconduct in some cases). It is stated that some authors are considered “hot-shots” and cited much more than other authors. Different mechanisms are proposed for this effect, ranging from writing a single, groundbreaking article and having spillover-effects from this, to cronyism and research experience leading to better articles. When compared to the survey variables, there is very high agreement on knowing an author influencing quality judgment of a paper prior to reading it, *goodauthor* ($P(v > 50) = .78, \bar{v} = 72.0$), indicating that participants explicitly connect authors to research quality. Also the experience of authors is seen as positively correlated with research quality, although more moderately, *authorexpr* ($P(v > 50) = .62, \bar{v} = 67.9$). Also when asked directly about whether authors are decisive for the quality of an article, *authorqual*, more survey participants agree than not ($P(v > 50) = .59, \bar{v} = 60.3$).

8.1.8. Factor 8 - Citation meaning

Index-scores for this factor are shown in figure 8.8.

Figure 8.8: Histograms of index scores for variables associated with factor 8 - Citation meaning



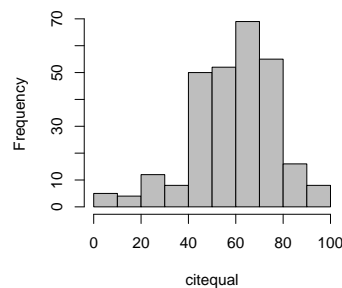
Several interviewees mentioned specifically that mainstream studies as well as studies on hot topics were far more cited than other studies, in which case the assumed connection between citations and research quality would not be present. Some of these comments were despondent, as they felt more specialised research was at a disadvantage. In opposition, others considered broad or mainstream research to be a positive thing, as it is potentially relevant to many (and therefore cited by many). Citations as a result of publishing on a hot topic was seen as a negative effect, referring to the concept as “fashion-studies”. In the survey, participants were asked to consider whether highly cited studies are an expression of high quality research (*citeequal*), recognition or collegial respect (*citerespect*) or the article being on a hot topic (*citetrend*). The participants agreed very strongly ($P(v > 50) = .91, \bar{v} = 80.2$) that trendiness has an

effect on citations. With regard to collegial respect, survey participants also agreed on an effect, although more moderately ($P(v > 50) = .67, \bar{v} = 62.9$). The question was based on some interviewees referring to “old guys citing each other”, and the increased attention to “hot-shot authors”.

8.1.9. Factor 9 - Citation quality

Index-scores for this factor are shown in figure 8.9.

Figure 8.9: Histograms of index scores for variables associated with factor 9 - Citation quality



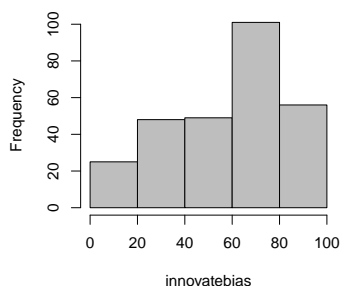
The connection between high research quality and a high degree of citedness, `citequal`, did not share a common factor with other variables. Survey participants agreed on the correlation ($P(v > 50) = .72, \bar{v} = 64.0$), although not as strongly as on the trendiness-effect, but more than the cronyism-effect (see factor 8). Interviewees explicitly stated that there is a connection between the quality of a research paper, and the citations it receives. This was also stated reversely; “if something is uncited after a period of five years, it is probably irrelevant”. Some interviewees found citations to be a more direct measure of quality than journal impact factors, but there was also an awareness that citations take time to generate and cannot be used as quickly. At the other end of the scale, one interviewee stated; “If a group of researchers becomes involved in a research area, they will be cited - that is not a quality stamp” (Interview M18). The views on the connection between citations and research quality were thus rather diverse, and as was found by Aksnes and Rip (2009), there may be an overly cynical perception of this connection.

8.1.10. Factor 10 - Innovation stunt

Index-scores for this factor are shown in figure 8.10.

The variable `innovatebias` was not derived directly from any statement, but merged from complementary statements. One interviewee stated, that low-impact journals were more likely to publish truly groundbreaking research, hinting that peer review may be part of the reason (see also under factor 1). Other interviewees stated that peer

Figure 8.10: Histogram of index scores for variable associated with factor 10 - Innovation stunt

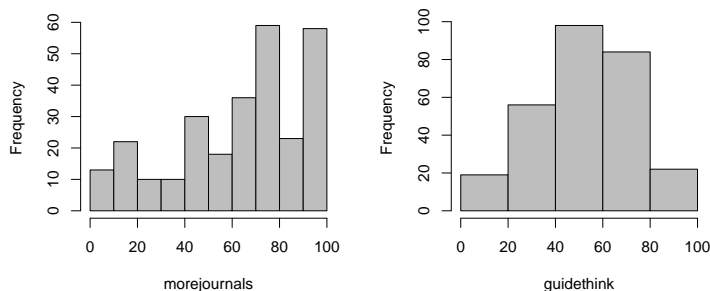


review was more strict in high impact journals, and these two effects were combined into one, claiming that peer review might prevent new and groundbreaking research from being published. The survey participants agreed moderately with considerable variance ($P(v > 50) = .56, \bar{v} = 60.3, Var(v) = 956$).

8.1.11. Factor 11 - Skepticism

Index-scores for this factor are shown in figure 8.11.

Figure 8.11: Histograms of index scores for variables associated with factor 11 - Skepticism



The majority of survey participants believed that more research journals than necessary exist, *morejournals* ($P(v > 50) = .7, \bar{v} = 67.6$). By the interviewees this was expressed both as an overflow of new, online journals as well as an overflow of research articles. The latter was especially stressed by many of the interviewees, claiming that it was not possible to keep up with the literature, that they needed to select articles based on e.g. citedness, specific journals or use by others. One participant stated that any research project should begin with performing a systematic review, to find out “what we don’t know yet”, and then do that study, because that participant believed the amount of redundant research to be too high. The same participant said, reproducing results should be limited to two or three studies, using better methodology than the original

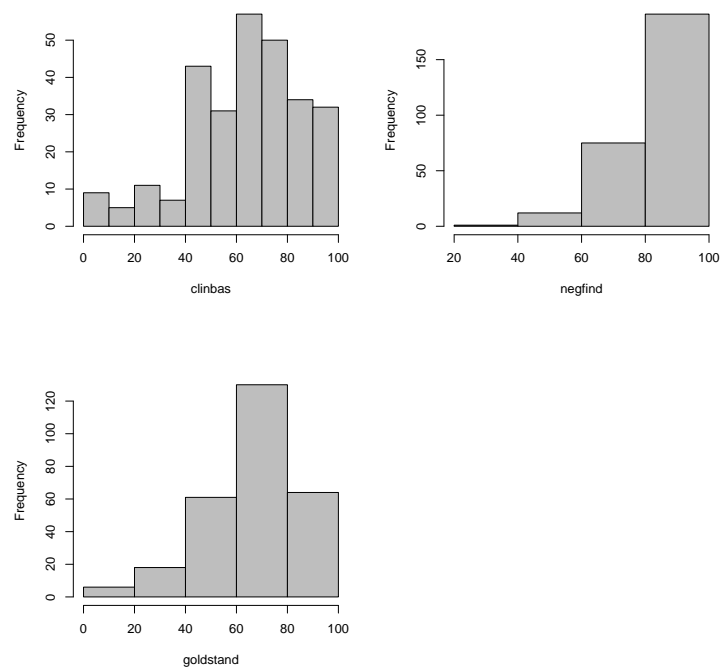
study - not 600 studies in two years, as expressed by the interviewee. The statement concludes with, “... you must continually ask yourself, »How can I contribute, as a researcher?»” (interview M18). In conclusion, the overflow of research articles is seen as a problem, and bolstering of resumés is suggested as a mechanism.

The `guidethink` variable is covered under factor 2.

8.1.12. Factor 12 - Propriety

Index-scores for this factor are shown in figure 8.12.

Figure 8.12: Histograms of index scores for variables associated with factor 12 - Propriety



The label “Propriety” for this factor relates to good conduct, and doing the right thing - a trait suggested by one interviewee as an indicator of research quality. One example of such good conduct was to conscientiously report any findings, correctly, detailed enough to reproduce, also in those cases where results were negative. Some other participants agreed on the importance of publishing negative findings, but they also expressed concern that it was difficult to publish these:

“...if you have half-bad data, and especially if it’s negative, they will never be published, right? Other than in Journal of Junk, and you can’t live off of that.” (Interview M08)

“But in reality, it’s difficult to publish something, if you have negative results. Even though it is just as valid, and has value for others to know how not to do it.” (Interview M05)

“... article which is reproducible, has methodology concerned with understanding, and the results - whether they’re positive or negative - that’s not it. But it is the result of the research that should be scientifically unambiguous.” (Interview M18)

The importance of negative findings is also reflected in the survey, with an almost unanimous agreement on the potential usefulness of negative findings ($P(v > 50) = .95, \bar{v} = 90.9$). Also the use of golden standards can be seen as a form of propriety. The closer details relating to this variable are reported under factor 5. The final variable in this factor, `clinbas`, is not as clearly related to propriety, as it is concerned with whether changes in clinical practice can be led back to advances in basic research. This was suggested by some interviewees as a connection between basic and applied medical research, but as was documented in factor 6, it was also suggested that the relationship was not always straightforward, and another participant suggested that while such a relationship existed, the path could be so long from the basic research to the actual clinical change, that it would be obscured. High survey agreement was observed for this variable ($P(v > 50) = .73, \bar{v} = 68.2$).

8.1.13. Summary

While not all factors should be treated equally with regard to importance - since some explain a greater part of variance than other³² - they have all been narrated here. If we take their importance into account, however, it becomes clear that the effects behind journal prestige, clinical guidelines and referencing behaviour explain the greatest part of the variance. This does not mean they are linked most to research quality however, but that they stand out the most. It is clear from the narrations that journals play an important role in the scholarly communication in medicine, and that the research community links them to research quality. Clinical practice guidelines play an important role in the same system as well, and may indicate policy or even health effects to the assessor, however, the perceptions of clinical practice guidelines are so diverse that they can not be used straightforward for research assessment purposes. Systematic reviews and meta-analyses might be considered as alternative, less direct, indicators of policy effects, as they represent the best, current evidence on a topic (under ideal circumstances). With regard to the referencing behaviour of medical researchers, the narration and variable scores reveal a mixed image to us, where persuasion and reward mechanisms in referencing appear to be in effect simultaneously. Both effects can be ascribed to each their version of research quality, however; a direct assessment of which dimension of research quality citations measure does not appear to be possible, as the references are

³²see previous chapter

not likely discernible. Further studies into the full-text context of references (e.g. Small, 1978, 1980) might reveal more about the nature of references, and their use in specific contexts.

A great deal of the other factors pertain to different stages or types of research, or to internal, latent attributes of research articles, such as the applied methods, clinical relevance and the interchange between basic and applied research. While each of these factors contributes with a smaller proportion of the total variance, together they provide a rather large proportion, and different combinations of these factors, along with the abovementioned will be included in the conceptual model described in the next chapter, as the just mentioned aspects of these factors are elements of research quality dimensions.

8.1.14. Assessment of participants' articles

As noted earlier, interviewees were asked to bring two or three articles as examples of high quality research to the interviews, and elaborate why they believed these to be high quality. In the following, we will first summarise the reasons interviewees gave for their selection of exactly these papers, which is followed by a comparison of the citations to the papers and an indicator of “expected” citations. For comparison, the summarised quality statements are listed in table 8.1, using a similar layout as table 8.2, containing the publication details and citations of the various articles.

Table 8.1: List of papers collected from main study participants, containing quality-related statements.

Paper ID	Quality statement
M01_1	Useful in a clinical setting.
M01_2	Useful in a clinical setting.
M03_1	Novelty of idea/hypothesis. Tested and tried.
M03_2	Novelty of method, good study design. Tested and tried.
M04_1	Good material, good methods. Not written to gain yet another publication, relevant and important message. Not salami-publishing. Multimodality.
M04_2	Article is a review: high degree of coverage. Thorough, has a message. Genuine/serious.
M05_1	Important problem, clinical relevance. Strong methodology.
M05_2	Important problem, clinical relevance. Strong methodology. Active reduction of bias.
M08_1	Importance for patients. Strong visualisation. Unique data.
M08_2	Long study time: 10 years work. International collaboration: large patient material, multicenter. Based on previous, basic research, leading to specific medicinal test.
M08_3	International and cross-disciplinary collaboration, large material, novel findings.
M09_1	(No specific responses, general remarks about novelty and importance of the hypothesis and methodology)
M09_2	Importance of hypothesis/problem. Elegant/creative use of data sources. Good article structure/reporting.

Continued on next page

Table 8.1 – continued from previous page

Paper ID	Quality statement
M10_1	Changed practice. Good journal.
M10_2	Review: all important topics included. Good author. Good journal.
M13_1	Well-structured approach. Theoretically based. Tested for flaws. “A significant step forward”.
M13_2	Basic research: Innovative, novel method. Basis for new studies.
M13_3	Basic research: Extreme detail, multimodality.
M15_1	Basic research: Tests causality rather than associations. Strong hypothesis. Adds new (basic) knowledge.
M15_2	Basic research: Tests causality rather than associations. Strong hypothesis. Adds new (basic) knowledge.
M16_1	Importance for clinical practice - changed practice.
M16_2	Review: Provides good overview, very useful in clinical practice.
M18_1	(General remarks about the necessity of conducting literature reviews prior to performing research, and to avoid conducting bad/redundant research)
M18_2	Review: Systematic and thorough.
M19_1	Review: Short. Good authors. Good overview. Good journal.
M19_2	Many tables. Good overview. Short/to the point.
M21_1	Very thorough (Double-blind RCT). Well-conducted, well-planned. Results strong enough to conclude no further studies needed. “How good research should be conducted”.
M21_2	Very thorough.
M25_1	Case report: Huge material from one case (author self). Covers entire disease progression from one case.
M25_2	Novel methodology, novel idea. Landmark-study. Establishes new research area (Connection between lung cancer and tobacco smoking).
M25_3	Strong design/methodology. Results strong enough to conclude no further studies needed.

In summary, the elements which interviewees considered high-quality were focused on the effects and internal aspects of the articles. Internal aspects are here meant to cover the methodology, hypothesis, research problem, results, materials and style of articles, while the effects pertain to the clinical relevance, health impact and academic impact thereof. There was direct opposition against what was referred to as salami-publishing, or writing articles purely for the sake of career advancement, which was considered low-quality research. Also novelty and creative approaches to solving problems were praised in several cases. A great number of these attributes can only be captured by peer review, however; towards the end of each interview, the participants were asked whether they were aware of the number of times these articles were cited. Only few were able to provide an answer, and in almost every case it was a question of guessing. As evaluative bibliometrics commonly employ citation counts as indicators of impact and by proxy research quality, these articles were compared to other articles. The references to the articles, as well as their citation counts and mean citations for papers of the same type, published the same year, in the same journal are listed in table 8.2.

8. Factor narration and conceptualisation

Table 8.2: List of papers collected from main study participants. PY = Publication year, TC = Times cited in Web of Science, total from publication until retrieval (09 Sep, 2011), MCJ = Mean citations for papers of type=article, published in the same journal, the same year with the same citation window as TC. References in appendix. One paper omitted as it could not be retrieved (M18_1).

Paper ID	Reference	PY	Journal	TC	MCJ
M01_1	(Rüschhoff et al., 2010)	2010	Virchows Arch	27	1.87
M01_2	(Mueller, 2008)	2008	Best Pract Res Clin Gastroenterol	11	7.32
M03_1	(Laurberg, Pedersen, Vestergaard, & Sigurdsson, 1991)	1991	J Int Med	169	29.03
M03_2	(Black et al., 1996)	1996	Lancet	1,828	119.97
M04_1	(Fischer et al., 2004)	2004	Mod Pathol	17	20.94
M04_2	(Slim & Mehio, 2007)	2007	Clin Genet	20	11.11
M05_1	(Hoberman et al., 2011)	2011	New Engl J Med	8	5.20
M05_2	(Tähtinen et al., 2011)	2011	New Engl J Med	6	5.20
M08_1	(Hvid-Jensen, Pedersen, Drewes, Sørensen, & Funch-Jensen, 2011)	2011	New Engl J Med	1	5.20
M08_2	(Olesen, Bouwense, Wilder-Smith, Van Goor, & Drewes, 2011)	2011	Gastroenterol	0	1.56
M08_3	(Neziri et al., 2011)	2011	Pain	0	0.67
M09_1	(Thomsen et al., 2009)	2009	Ann Intern Med	11	31.46
M09_2	(Lohse et al., 2007)	2007	Ann Intern Med	172	55.31
M10_1	(The GRIT study group, 2004)	2004	Lancet	75	128.66
M10_2	(Baschat, 2011)	2011	Ultrasound Obstet Gynecol	1	0.55
M13_1	(Mayberg et al., 2005)	2005	Neuron	656	84.41
M13_2	(Livet et al., 2007)	2007	Nature	236	123.26
M13_3	(Koutcherov, Mai, Ashwell, & Paxinos, 2000)	2000	J Compar Neurol	30	36.00
M15_1	(Podrez et al., 2007)	2007	Nat Med	85	108.34
M15_2	(Febbraio, Guy, & Silverstein, 2004)	2004	Arterioscler Thromb Vasc Biol	57	39.89
M16_1	(Rivers et al., 2001)	2001	New Engl J Med	2,432	305.21
M16_2	(Eissa, Carton, & Buggy, 2010)	2010	Br J Anaesth	4	2.61
M18_2	(Zitvogel, Kepp, & Kroemer, 2011)	2011	Nat Rev Clin Oncol	2	1.42
M19_1	(Laurent & Boutouyrie, 2007)	2007	Hypertens	60	26.00
M19_2	(Gedikli et al., 2010)	2010	Clin Exp Hypertens	5	0.87
M21_1	(Liaw et al., 2004)	2004	New Engl J Med	701	320.03
M21_2	(Wiegand et al., 2008)	2008	Antivir Ther	19	10.56
M25_1	(Klatskin & Kimberg, 1969)	1969	New Engl J Med	259	90.04
M25_2	(Doll & Hill, 1950)	1950	Br Med J	787	10.18

Continued on next page

Figure 8.13: Relationship between citations to selected articles and comparable articles. TC = times cited for selected articles, MCJ = Mean citations to articles of the same type, published the same year in the same journal.

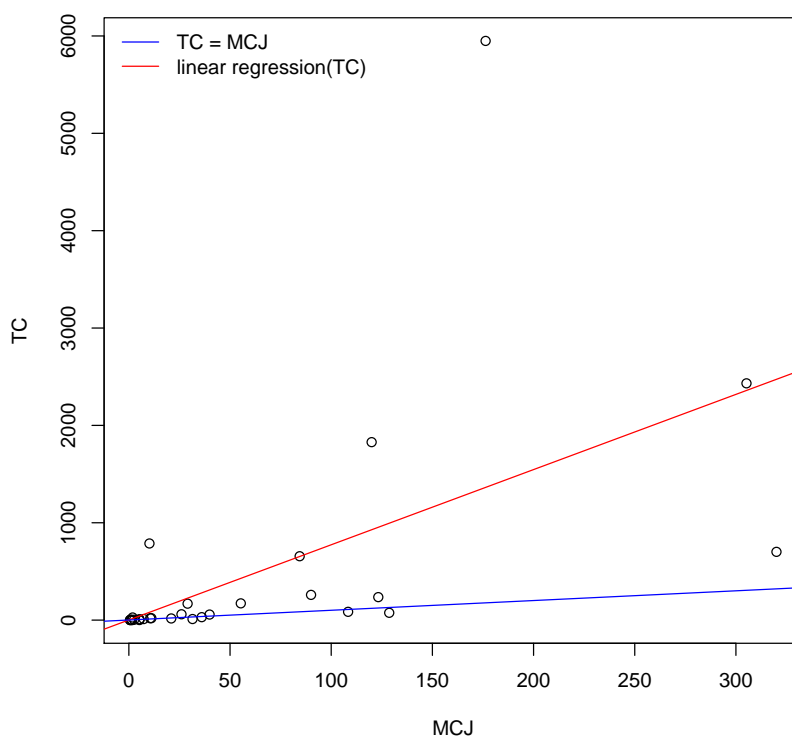


Table 8.2 – continued from previous page

Paper ID	Reference	PY	Journal	TC	MCJ
M25_3	(Scandinavian Simvastatin Survival Study Group, 1994)	1994	Lancet	5,949	176.31

To illustrate the degree of citedness more clearly, the citations per article as a function of mean citedness for the journals they were published in are plotted in fig. 8.13. The individual articles are shown as circles, while the red line is a linear regression fit for the data ($y = 7.7x + 1.4$, $R^2 = .31$) and the blue line represents the boundary at which citedness for articles equals the mean citedness of journals. As can be seen, the selected articles are cited almost eightfold of what could be expected.

While this relationship in no way means citedness and research quality are equivalent - and especially not based on so relatively small a sample - these findings provide evidence about some type of correlation between the most highly cited articles in a journal and an ad hoc use of the concept research quality. This observation falls in line with normative theories of citations, and also matches our observations from the survey concerned with citations and research quality. In addition the sum of qualitative

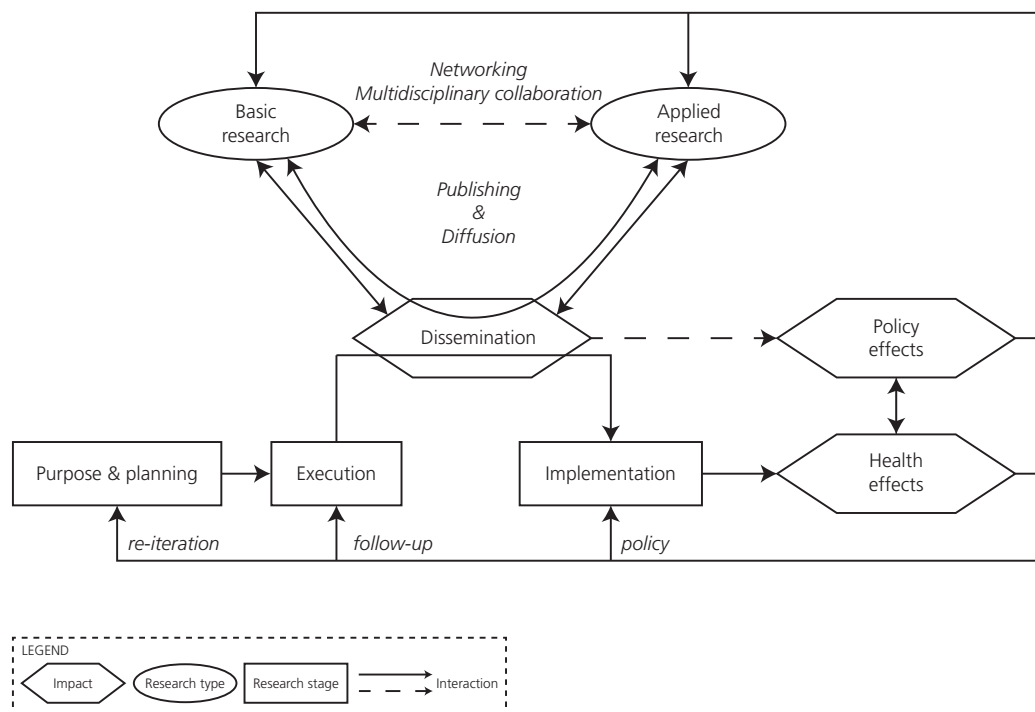
attributes, which would normally be assessed through peer review appear to correlate well with high-citedness.

8.2. Conceptual model

Three different questions related to the concept research quality were asked initially in this dissertation: How is research quality perceived in medicine; how do the identified aspects of research quality relate to each other; and how do bibliometric methods apply to the quantifiable aspects of research quality? In the course of the previous chapters, data, results and analyses have been put forth to provide the background for answering these questions. The first question has been answered by identifying and narrating common factors related to research quality in medical research, and the second question will be answered by collecting these factors and variables in a conceptual model during the following section. Finally, we will relate this model to bibliometric methodology, in order to address the third and final question. When reading the following sections, it should be kept in mind that the grand perspective of this dissertation is evaluative bibliometrics, and the focus of the conceptual model is on the aspects of research related hereto, as is also argued in section 3.1. While other elements may be mentioned for the sake of completeness, the reader should not expect them to be elaborated.

Central to bibliometrics, and the entire research world, lies dissemination; the act of spreading knowledge through various scientific and scholarly channels, and especially through publishing of research results in scientific journals, books, conference proceedings and various other media. Dissemination is thus a central element in our model, and can similarly be found on a prominent position in the model of Bazeley (2010). Another element, which has been mentioned by almost every interviewee, and is also used as an endpoint in the model of Lewison (2004), is the effect of medical research on general health. Lewison proposed a number of steps on the way to better clinical care and less illness incidence, including government policy, clinical guidelines and medical education (see fig. 4.1, p. 56). In our conceptual model, we will include this dimension of research, but as it is not as strongly related to bibliometrics, as the dissemination is, we will limit it to two major effects; health effects and policy effects. Health effects are all those direct changes in general health, caused by research in one way or another, while policy effects are e.g. guidelines for clinical practice, but also local, national and global political decisions on funding or treatment. These three areas, dissemination, health effects and policy effects, are considered impact areas in our model. There is a duality in the use of impact to describe these areas, as it is used in the sense of research having an impact on these areas, but also that this is what has an impact on the world. Finally, if we wish to measure the impact of research, it is in these areas. However, several quality-related factors influence whether research has an impact or not, and perhaps it is possible to not just measure the impact, but also some of the underlying mechanisms. To illustrate this, a visual model of the concepts is shown in fig. 8.14. The model consists of three main dimensions, labeled impact, research type

Figure 8.14: Conceptual model of research quality in medical research, from an evaluative bibliometrics perspective.



and research stage. It revolves around the impact dimension which interacts with the other dimensions. Below is a broad description of the model, which is then followed by a more detailed one, including the identified factors and their roles in the model.

The two research types, basic and applied research, are included to represent the interactions between them, as this interaction was a point stressed by several participants, as an aspect of research quality. The types primarily interact by means of dissemination, but multidisciplinary collaboration also enables the transportation of ideas and knowledge from one area to another. The other dimension of the model, research stage, can be regarded as a subdimension of the research types, but as there are several quality-aspects to be found in this dimension, it is illustrated in its own right here. Due to the subdimensionality, however; no interactions are drawn directly between research types and research stages.

The initial stage of the research process is planning, and finding a purpose of the research. Essential attributes of this element are funding, and the relevance of hypotheses; is the research project asking relevant questions, and can the results be expected to lead to impact? The planning aspect also involves considerations about methodology, the number of patients to include, the study design and ethical considerations about the potential harm done, compared to the expected benefits. The execution stage, i.e. the actual conduction of the research project, is not very interesting from a bibliometric perspective, as it is mostly beyond the grasp of evaluation. It is primarily the methods

used in the research that are of any use, and it is only when the research is published that this is accessible. Once the research is published, however, this stage can be evaluated as well, and it is possible to see, if the research becomes directly implemented, the final stage of the research translation process, or if it affects policies, e.g. in the shape of new clinical practice guidelines, which in turn may result in practice implementation and hopefully a health effect. The feedback arrows from health and policy effects to the research types and stages symbolise the iterative progression of research. These effects can be very direct, e.g. in new policies leading to improved funding of specific research areas, or changes in health status leading to follow-up studies of specific treatments. The feedback to research types is generalised, while the feedback to research stages illustrates the levels on which the feedback is most relevant.

With regard to the identified factors, they all affect different parts of the model, and each part of the model can be related to one or more factors. To begin with, the interaction and relationship between basic and applied research is described by factors 6 and 12. Concerning research quality this means that an exchange between the two types - measurable as references from e.g. applied research to predecesing basic research - can be used as a measure of diffusion from one area to another. This should not be confused with earlier work on citation diffusion (Frandsen, 2004; Frandsen, Rousseau, & Rowland, 2006), as the current use of the term diffusion focuses on the exchange between basic and applied research. A ratio between the citations from either area could thus give an indication of whether a basic research unit was succesful in disseminating knowledge to the applied fields. This will be discussed further in section 8.2.3. The interaction between either research type and dissemination is explained by the factors 3, 8 and 9, all relating to citations. From the narrated factors, and the variable scores we have learned, that there is some trust in citations as an indicator of research quality, but we have also learned that medical researchers associate citations with other mechanisms than research quality. We have also seen how the papers they selected as examples of high quality research were cited somewhat better than average, and it appears there is some connection between high citation counts and the quality of research articles, although it is not straightforward. That being said, there are several indications - also outside of this dissertation³³ - that citedness is related to research quality, that it might be used as a generalised proxy thereof, and that the connection is stronger, when the articles are cited more. However, if we analyse the remaining factors with regard to the conceptual model, we will see that citation, reference and document analyses can be used for more detailed and specific bibliometric evaluations of research quality aspects as well. This will be discussed further (see section 8.2.3) after the remaining presentation of the model.

The initial research stage, purpose and planning, is explained somewhat by factor 5, with some additions from factor 1. Factor 1, journal prestige, is included here, as journal impact factors were considered important for grant applications. While this

³³See the review and theory chapters, in particular with regard to the normative approach to citation theory.

practice can be considered problematic, it is a factual element of the funding process, and something medical researchers need to take into consideration when publishing their research. Whether there is a connection between funding and research quality is pure speculation at this point, and far from the focus of this dissertation, but one could easily imagine an overall correlation between the ability of a research group to procure funding and the overall quality of their research. The other elements of this research stage, relating to factor 5, are primarily concerned with the purpose of the research, of asking research questions that matter, and planning the research project appropriately in accordance with the posed questions. This becomes even more important in the following stage, the actual execution of research, which is explained by factors 4 and 12 in addition to factor 5. Also a great number of the article-level variables can be used here to illustrate what is perceived as high quality research, as several of the highly selected values focus on a strong methodology, documenting this for the sake of reproducibility and having a good connection between the hypothesis and the applied methodology. Bias and validity are considered even more important, and so is the quality and selection of materials. These variables naturally don't become available before the research is actually disseminated, however, they relate to this stage of the research process. Whether it is viable or even possible at all to assess these variables bibliometrically is questionable though, but there might be a potential for automated content and context analysis.

The implementation stage of the research process does not follow as directly from the previous stage, as the execution does from the planning, and in some cases the execution and implementation are entangled (e.g. in intervention studies). The factors most closely related to the implementation are factors 2 and 11, which are both related to clinical guidelines which are precisely used to implement research in practice. This implementation can either be direct, as illustrated in the model, which is most likely to take place in clinical hospital departments conducting research locally, or by being adopted in a clinical practice guideline regionally, nationally or even internationally. In this case, the research must first have a policy effect, which can in turn lead to an implementation of research results. This type of implementation has the potential for a much larger health effect, but as can be seen from the survey and interview results, there is large variation in the trust that participants hold in clinical guidelines.

The various stages in the research process, as illustrated in the conceptual model, can be granulated much further and feedback loops between various stages are present as well. Several such models exist, however; this is not the focus of the current model and thus we only use the three broad stages displayed in the model. We also only use one-way interaction, although some feedback is assumed, as it is the progression from one stage to the next that is most interesting for bibliometric research assessment. As an example, it is invisible to bibliometrics whether an article has followed a direct path from funding to publication, via some planning and execution, or if twenty iterations on more detailed steps on the path have been made. Once the research is published, however, it might be possible to track new iterations based on the published results, which is the reason for including feedback arrows in the model.

The factors 1, 3, 4, 5, 7, 8, 9, 10, 11 and 12 are all related to dissemination in one way or another. They cover journal and author related mechanisms, peer review, and other publishing-related variables. Most obviously related to research quality are peer review and journal prestige (which participants also considered to be closely related), but also the credibility and experience of authors were suggested to be related to research quality. These are all traditional bibliometric variables, which we might have expected to find in such a model, as is the more subjective peer review. The interesting aspect is to see the dissemination in relation to the other elements, and consider the interactions and paths towards impact, and this is where this dissertation provides novel insights, in addition to articulating the different aspects of research quality. The role of citations in this context is briefly discussed below in section 8.2.1 and the specific role of the Journal Impact Factor in health research (in particular with regard to funding) is discussed in section 8.2.2, while the potentially new applications of bibliometrics will be treated in two sections on citation paths (section 8.2.3) and health and policy effects (section 8.2.4).

If one compares this model of the research process, and the research quality aspects, to the previously mentioned models of Björk (2007) and Lewison (2004), it should be obvious that there are central resemblances, but also significant differences, especially with regard to where quality can be found and measured. The focus of the other two models are quite different, in the way that Björks model is on the research process, and that of Lewison is also on the research process, but involving numerous indirect and direct influences. The model presented here includes the research process, and from that perspective resembles a simplified version of Lewisons. The research quality aspects, which are mainly presented textually, are additions to the previous work, but can be compared to the findings of Bazeley (2010). Again there are both resemblances and differences, and both can be considered significant. Several of the elements present in Bazeleys model are also identified in the model of this dissertation; dissemination plays a central role in both models, and Bazeley also includes variables relating to research practice, such as methodological soundness, technical skill, substantive knowledge and strategic skill. Also collegial engagement can be found in both models, although presented somewhat differently. Where Bazeley focuses on the interpersonal skills, the participants in this study focused more on professional networking (in Bazeleys model seen as part of dissemination) and on the meaning of being a “big-name” in medical research. Bazeleys model of research performance consists of several elements, but they are grouped into two main categories, research activity and performing, where the research activity category focuses on interior values and activities, while “performing” focuses on the exterior (dissemination and collegial engagement). In other words, the resemblances between the model presented here and Bazeleys can be found on the exterior dimension, which is elaborated further in the present model. While we can learn something from both models, the present model appears superior for bibliometric assessment of medical research, as it is more closely related to both areas, than Bazeleys model (which is rather a universal, general model). Compared to the aforementioned

three models we can thus find several resemblances from different perspectives, but we can also see novel perspectives in the present model. The unique part is the combination of different impact dimensions with both the research process and the research types, as well as the interactions between them, and the quality attributes of the interactions and individual elements. With regard to impact, the distinction between different types of impact falls in line with the recommendations of Kuruvilla et al. (2006), although it is not as finely granulated here. In conclusion, the present model validates previous findings and combines central elements from different aspects of research performance assessment, resulting in a specialised, conceptual model of research quality in medical research, with special attention to evaluative bibliometric measurement. In the following sections, we will discuss the implications of some of the observations made in this regard.

8.2.1. Citation impact

As was shown in the initial chapters of this dissertation it is not new to observe a relationship between citations and research quality. What this study adds is an articulation of the different roles citations may play in that specific context. While numerous other studies and handbooks provide insight into the role of citations in academic research, it has not been possible to find other models or descriptions visualising the roles specifically with regard to research quality. This role is centered in the dissemination impact area and can be seen both as a process and an effect. The process is described in section 8.2.3, while the effect is equivalent to what is traditionally coined “impact” in evaluative bibliometrics. The number of citations a given article receives can thus be interpreted as the ability or success of dissemination - if any article is highly cited then that article successfully manages to spread results, data, knowledge, information, ideas, hypotheses etc. to a larger audience. This can then be further broken down: how far does this impact spread? To which fields? How successful are subsequent articles, using the current as reference? But it can also be used as a stand-alone indicator of impact: how often are researchers from a given university or research group cited? High numbers (with proper normalisation steps) tell us researchers from a given group are able to disseminate their research better than those with lower scores. But is dissemination a quality? Or is it merely a prerequisite? Considering the central role dissemination plays in the progression of research

We know from the literature, but also from the responses provided to the interviews and surveys, that citations are affected by a number of different factors, such as timeliness, study design, correct use of methods, references and rigorous reporting, publishing journal, the prestige of (co-)authors, number of pages, availability, collaboration etc. (See sections 3.2.2, 3.2.3, 6.3 and 7.2, but also Lokker et al. (2008); Eysenbach (2006); Callaham, Wears, and Weber (2002); White (2004) & van Raan (1998a)). When citations are aggregated it becomes impossible to discern these different causes from each other, but this is a feature or advantage rather than a problem. Each of these effects can be studied individually, if so desired, but the combined citation score offers

us a construction of a new whole. Again, this is in no way a new observation, but rather than using a highly ambiguous term such as “impact” on its own, by qualifying it as dissemination-success we can separate this type of academic impact from other types. It is not our intention to change the use of a fine term in a specific discursive community, however, it might be useful to use qualifying terms, such as “dissemination impact” or perhaps “academic dissemination impact” in broader communication, signaling the precise meaning of this term, the role it plays in the research system and as a dimension of research quality.

A different type of impact indicators, altmetrics (Priem, Taraborelli, Groth, & Neylon, 2010; Priem et al., 2012), measuring the impact on academic bookmarking sites as well as social media might also benefit from this elucidation of the impact concept. One of the aims of altmetrics is to provide a more complete image of impact than citation-based indicators can on their own (Priem et al., 2010). While there are numerous problems with the validity and interpretation of altmetrics (Wouters & Costas, 2012; Rasmussen & Andersen, 2013) there is also potential (e.g. Priem et al., 2012; Piwowar & Priem, 2013), if kept in mind that altmetrics and bibliometrics are complementary. One step towards this might be to clarify which types of disseminative impact altmetrics analyse, and how this relates to citation-based academic dissemination impact.

With regard to citation impact as a dimension of research quality, we have found that while researchers believe in citations as representative of the quality of a research article, they also believe (and more strongly) that factors such as timeliness influence citation counts more. The different effects found in this study, of which some of them are confirmed elsewhere (see above references to other sections), can be summarised into two impact-components: a communication component and a latent component. The communication component covers effects such as timeliness, language skill, style, and similar, that let the author convince the reader about the impact of their work, while the latent component is the underlying research, the strength of methods, meaning of results, representativity, validity and originality. It is reasonable to assume that long-lasting citation impact will require at least some of both components to be present. This model explains cases such as sleeping beauties (van Raan, 2004) as a low communication-component (e.g. lack of timeliness). This construction of impact also stresses an important aspect of contemporary research: if a study is not communicated (well), it might as well not have been executed, as there will be no benefit to anyone beyond the original researcher. By representing both of these dimensions of research quality in one construct, citation-based impact is confirmed as a very powerful and useful indicator.

8.2.2. Journal Impact Factor

The journal impact factor is quite likely the best known, most used and also most polemic of all bibliometric indicators. Several studies have beyond doubt shown several technical and interpretational problems with the indicator (e.g. Seglen, 1992, 1997; van Leeuwen & Moed, 2005; Leydesdorff & Bornmann, 2011), yet it is advertised by

publishers and used in grant application processes (see Moed & van Leeuwen, 1996), likely more than any other indicator. In some connections the journal impact factor has been interpreted as an “expected impact” of articles, although it is quite clear from the aforementioned studies that the connection between the journal impact factor and the citations to individual papers is very small. Thus there seems to be a mismatch between what bibliometricians know - the journal impact factor is flawed and misrepresentative of individual papers - and what individual researchers might think - that the journal impact factor is something to strive for. But perhaps the journal impact factor has become a de facto quality indicator through its use rather than its technical merits. When a grant committee uses journal impact factors to assess grant proposals, it becomes important for researchers to achieve high scores - and thus they are more likely to aim their research at high impact journals, which thus becomes a self-feeding mechanism, as the journals are able to choose the best research out of a larger amount of studies. What we should learn from this is that there is a need for a bibliometric indicator that authors can use relatively easily, when submitting grant proposals and similar, and that medical researchers have very high trust in journals as indicators of article quality (see variable `journalquality`, in figure 8.1). It therefore makes sense to have a journal-based indicator of academic impact in medical research, despite the bibliometric concerns this raises, however; as long as the research system uses impact and journal prestige as a part of it - as a language or means of communication, as was stated by one interviewee (see section 8.1.1) - it is doubtful that the bibliometric research community can prevent such behaviour - we can only strive to provide better indicators and make them available for the research community.

With regard to the journal impact factor as a de facto quality indicator, this is a problematic construct. The indicator is removed from the actual quality of the individual article, but it is used as an indicator of precisely this. We can not ignore this fact, and we must accept publishing channels as something a great number of researchers believe in as a dimension of research quality. Based on the interviews, however; we also know that researchers are not necessarily happy with this system (see section 8.1.1), and there is thus potential for a change in the current system - especially as those expressing the greatest lack of confidence in the current system were those at a disadvantage, e.g. from research areas with low overall journal impact factors. This observation falls in line with a correlation (or lack of) between researchers’ perception of journal esteem and citation-based journal indicators (Lewison, 2002b), with a higher correlation for “hard” medical research fields than “soft” healthcare research (e.g. nursing).

8.2.3. Citation paths

It was briefly mentioned above, that research quality might be related to citations in other ways than citation counts, e.g. in the shape of diffusion. The term is here used for two different types of citation exchange between different elements. On the highest level, the diffusion occurs between basic and applied research, and on the lowest between different study types. Other types of citation diffusion exist as well, e.g. be-

tween research groups, journals etc. A large body of research on these types of citation diffusion already exists, and while it is relevant, the observation is not new, and will not be discussed further here. Nor can it be claimed that the idea of knowledge/idea transfer between basic and applied research is new. In fact it is an essential thought behind the Triple Helix of university, industry and government innovation (Leydesdorff & Etzkowitz, 1996; Etzkowitz, 2003). But where the Triple Helix focuses on converting university research (which is often basic research) to innovation (sometimes used as synonymous with certain types of applied research), e.g. in industry, the meaning of diffusion in the current context is another. Diffusion is here meant to signify the adaptation of research from one field into another. To relate it directly to medical research, this could be a case where a basic research group discovers a new genetic marker, which enables the targeted testing and diagnosis of specific cancer types. In this case, the latter is dependant on the basic research (discovering new marker) being executed and published. In other cases, it was suggested by one interviewee (M25), it may be the other way around - clinical case reports of certain disease stories, extraordinary treatments or similar may lead to new hypotheses, sparking basic research in a number of areas. A famous example of this was the case report of Richard Doll, published in the *British Medical Journal* in 1950, leading to the first suspicions of tobacco smoking as a cause for lung cancer. In other words, and as also depicted in the model, the diffusion may go in both directions, and either type of impact is just as important. Another aspect of this diffusion is discussed by Shaw and Chase (2007). Among other, they argue for a new, more cost-efficient approach to medical research, focusing more on basic research and less on clinical trials. They compare to engineering, where the main research effort is in basic research - which is considerably cheaper to conduct - leading to few, but very focused, applications with a high success-rate, while they claim the success-rate of clinical trials to be much lower and more costly (Shaw & Chase, 2007). An efficient diffusion between these research types might thus lead to improved success-rates, which clearly is a dimension of research quality. With regard to bibliometric measurement, the use of diffusion metrics may provide novel insights into the differences between certain research groups, or even universities. In classical university evaluations and rankings, such as the Leiden ranking³⁴, the main (bibliometric) indicators are based upon citation aggregates in one form or another, and universities are ranked according to their overall impact. If we were to supplement such indicators with metrics showing the proportion of citation diffusion between basic and applied research for each university, we would gain even more information about the research production of these universities, namely about their skill in leading basic research towards practical application, or the creative ability to produce new hypotheses. These are important, research-quality-related skills, which can not be measured through citation counts, but by measuring the citation diffusion, we might grasp a glimpse of it. Creating an index or ratio of diffusion would also allow us to test the hypothesis of whether there is a connection between the degree of clinical trials' foundation on basic research and their success rate.

³⁴www.leidenranking.com

While the abovementioned can be seen as a straightforward “adaptation indicator”, it might also prove to be too simple. As was mentioned by one interviewee, the progression is quite likely more complex than this. Rather than just looking at the ratio of citations basic research receives from applied, more information could be obtained by following citation paths. The idea of citation pathways was suggested by Small (2000), albeit as a discovery tool, while we will here suggest using it as a type of impact or adaptation indicator. By citation paths we mean following the consecutive citations, so when paper *A* is cited by paper *B*, which is in turn cited by paper *C*, a citation path is tread from *A* to *C*. If all papers A_i are basic research papers and all papers B_j are applied research papers, we can use any sequence of $A_1 \leftarrow A_i \leftarrow B_j$ within a given timespan, or within a given graph-theoretic distance to give a more precise ratio of citations from applied to basic research, allowing for iterations on the basic research before reaching an applied research endpoint. This approach would allow for an in-depth analysis of the adaptation of basic research, encouraging the more cost-efficient research approach suggested by Shaw and Chase (2007).

8.2.4. Health and policy effects

Another type of diffusion, or citation paths, to be discussed here pertains to the different medical study designs which most clinical research is based upon. While we cannot bibliometrically measure the health effects of individual studies, by looking at citation paths we might gain a glimpse, especially by looking at phase III and IV clinical trials and clinical practice guidelines as final destinations of the citation paths.

Phase III and IV clinical trials are specific study types for drug testing on large populations and following the marketing of drugs respectively. These studies are thus conducted after a given drug has proven to have an effect, and it is therefore reasonable to assume that the phase I and II trials have had an impact and relevance, since the trials are continued - and this should eventually lead to a health effect, although the size and type of this effect is not necessarily measurable. Also the clinical practice guidelines (CPG) can tell us about a different type of impact than the academic, citation-based one. The inclusion of a reference in a clinical practice guideline is equivalent to this study having a policy effect, as the guideline is a type of policy. However, there is large variation in the usage, quality and perception of CPGs (e.g. Ferket et al., 2010; van den Berg, Engelhardt, Haanstra, Langius, & van Tulder, 2012; Andersen, 2013), and such an analysis is therefore far from straightforward. If however we are able to identify successful CPGs, we have a positive indicator of both policy and health impact, which can be traced back through the literature through references. An alternative to using CPGs could be high-quality systematic reviews and meta-analyses as specific document types indicating policy impact. While these document types are not a priori policy documents, they do represent the highest level of evidence available, and as such are likely to be used by clinical practitioners.

More informative studies could be conducted in a collaboration between e.g. clinical epidemiologists and bibliometricians, as the effect size of the health impact in some

cases can be estimated through epidemiological studies, which could in turn be combined with citation path analyses. An analysis of the actual health impact of end-of-the-path documents, such as the abovementioned, would provide us with a connection between health impact and the traces back to the research leading to these changes. A very preliminary step into such comparisons was made by Andersen (2013) as a result of the discussion presented here.

There are many details about citation paths which need to be studied further, but it is an area which potentially can provide much new insight into the citation-based impact and thus a fruitful area for further studies.

8.3. Summary

Through the construction of a conceptual model, involving the identified dimensions of research quality, their relationship and their relationships to the overall research system, we have addressed the second research question and focused the reply to the first research question by classifying the various variables and factors into three main dimensions: dissemination, policy effects and health effects. A discussion about bibliometric methods in relation to these dimensions has been initiated, with the aim of reaching a final conclusion in the following chapter. In order to summarise, however; the main findings are as follows. Traditional citation analysis has been articulated as an analysis of academic citation impact, in order to discern this type of impact from numerous other impact types. Different mechanisms behind citation impact have been discussed, based on the identified factors, culminating in a two-component object. The two components, communication and internal qualities, relate to both good and effective communication about research results, e.g. by publishing in the right journal, and through well-written, convincing text and imagery, as well as the underlying research behind the actual publications, such as methodology, involved patients, the hypothesis etc. Specifically the journal is used as a special case, as the survey revealed a very high trust in journal esteem, and it would appear to be a very effective communication tool. The use, development and mediation of bibliometric journal indicators thus appears to be an important task for bibliometricians, in particular for medical research.

In addition to citation impact, the diffusion of knowledge between basic and applied research is also seen as an area with potential for citation studies to reveal information about the quality of research. This is exemplified by a citation diffusion factor, which could indicate e.g. to what degree the publications of a basic research group becomes adapted by the applied research communities. More detailed analyses about the citation paths are also suggested as a more complete, but also more complex, method for investigating diffusion.

The final themes to be discussed in this chapter were the policy and health effects. While especially the latter may appear difficult to measure bibliometrically, suggestions are made for citation path approaches for both effects. Policy effects may be measured as the inclusion of references in e.g. clinical practice guidelines, but also systematic reviews

and meta-analyses might indicate policy effects. Health effects are preferably investigated in collaboration with domain experts, e.g. clinical epidemiologists, specialising in measuring the health state of populations. However, similar reference inclusion, but in phase III and IV clinical trials, might also be used for bibliometric analyses of health effects.

The combined results of the model presented here, and the discussion about the relationship to bibliometric methods will be used in the following chapter to first give complete responses to the research questions, and then to present a model focused more on the dimensions of research quality, and less on the contexts as was the case in the conceptual model presented in this chapter.

9. Discussion

Throughout this dissertation an attempt has been made at drawing in the original research questions in relevant contexts and providing small, specific summaries of the key observations and assumptions of chapters or sections. Nonetheless not everything has been possible to capture in these summaries, and the research questions have not been treated together in one combined text block. That is why this discussion chapter will begin by listing the three research questions and summarising the findings from the previous chapters in relation to the questions. Thereby we hope to provide a clear overview of the exact findings of this dissertation. As the most important findings have already been discussed somewhat in the previous chapters, the focus of this chapter will be on relating the findings to other research and the potential implications of the results.

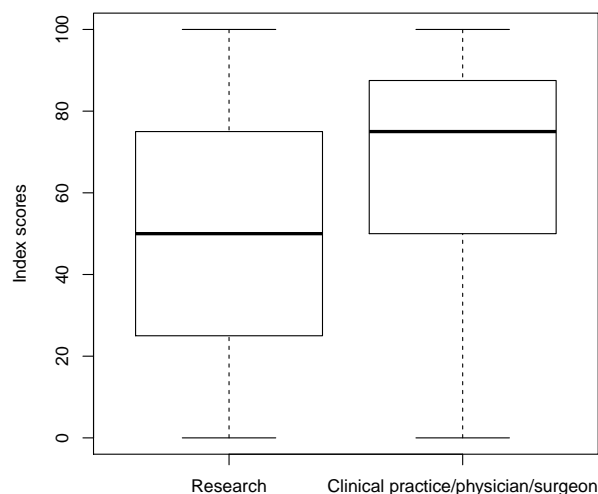
9.1. Research question summaries

How is research quality perceived in medical research? Through the interviews we have gained an impression of what participants thought to be the most important dimensions of research quality, and this impression was summarised into a number of statements. This process must necessarily be subjective, no matter how much effort is invested in diminishing bias, and while the results do reflect the views of the interviewees, we must be aware of the (involuntary) filter applied by the author in compiling statements.

The uncovered statements focus on a number of different topics which are clearly related to research quality, and represent what we have called dimensions of research quality. In the previous chapter, a conceptual model was built on these statements and quantitative expressions of them; the three impact types identified in this model are all heavily related to the most commonly used interview statements. The health and policy effect of medical research was heavily emphasised, and used in several article-descriptions, often as the ultimate goal and purpose of medical research, and also as

a means of deciding which research matters. To many of the participants the term impact, without prior introduction, was equivalent to clinical impact, but also to journal impact factors. With regard to the clinical impact, methodology and the use of golden standards (e.g. randomised controlled trials) were regarded as highly important aspects of quality, according to survey variables, but also the clinical relevance of studies was regarded as important. There was some difference in the reactions to clinical relevance in the survey, with scores distributed almost evenly over the entire range of response values. It could be suspected that this could be dependant on the type of participant, as clinical practitioners would most likely stress this more. This is somewhat confirmed by regarding the scores of the `clinrel` variable as a function of primary job category, as shown in fig. 9.1 (only showing the research and clinical practice respondents), clearly illustrating how clinical practitioners rate the importance of clinical relevance in all research much higher (median = 75) than those primarily concerned with research (median = 50). While health impact is thus considered an important part of research quality, the views vary depending on job function and possibly research areas, as clinical and biomedical researchers are likely to prioritise clinical relevance differently as well. From a qualitative viewpoint, clinical relevance was also related to the different stages of research, in that asking the right questions, testing the right hypotheses, including the right patients and using the right methods were also seen as elements of clinical relevance - but as this was not specifically included in the survey, we can not expect all participants to have reflected on these aspects of clinical relevance.

Figure 9.1: `clinrel` variable as function of primary job categories "research" or "clinical practice".



The dissemination of research was also considered an extremely important dimen-

sion of research quality, and several factors were identified within this topic. The variables which resulted in the most positive responses were those relating to peer review and journal esteem, equivalent to factor 5 (subjective quality) and 1 (journal prestige) respectively. The agreement on journal impact factors as indicative of the quality of individual articles was not as high as that for journal esteem. In other words, medical researchers appear to trust in journals as quality indicators, but not necessarily in bibliometric journal indicators. There was very high trust in peer review, although it was also seen as a gatekeeper rather than a high-quality indicator, and some interviewees also pointed to flaws of the peer reviewing system. The factor 5 combines these peer review variables with the golden standards and clinical relevance variables representing the other impact aspect of medical research. The policy impact is best represented by factor 2, which is concerned with clinical practice guidelines. It is symptomatic of the perception of clinical practice guidelines that the scores are very moderate, mostly centered around the midvalues, signaling that participants did not consider this aspect to be as important as the other, and further research may be required into the bibliometric and research quality related aspects of clinical practice guidelines.

In summary, medical research quality can be summarised into three multifaceted, interrelated impact types, each of which can be broken down into quantifiable elements, which will be discussed more under the third research question.

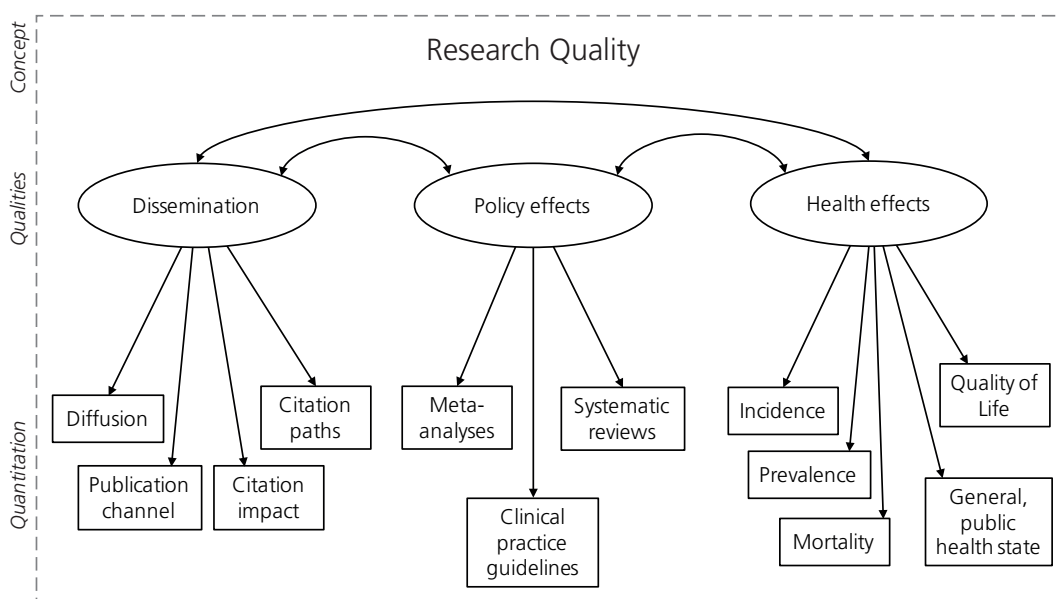
How do the identified aspects of research quality relate to each other? It was first shown how the identified variables were related, using factor analysis, and subsequently a conceptual model was built from these factors, showing a broader relationship between the abovementioned impact types and the types and stages of research. The identified research quality dimensions are mutually dependant, in that dissemination might lead to policy effects, which in turn may lead to an implementation of research and subsequently health effects. While there is a natural order to these types of impact, there is also the option for feeding back into the system, and actual health impact may in turn lead to new studies which are then disseminated and may in turn lead to further policy or health effects. The policy effects are mostly to be seen as a stepping stone from dissemination towards health effect, however, one that might be measurable bibliometrically, and which signals a potential health impact. Health effects, as has been discussed previously, are not likely to be analysable using bibliometric methods, however, collaboration between clinical epidemiologists and bibliometricians can potentially add new insights into the health effect dimension of research quality. The consequences of these relationships with regard to bibliometric evaluation will be discussed more intricately in response to the final research question.

It is important to understand that the model somewhat represents an ideal world in that the number of iterations, loops and detours are not shown, and the reader might be lead to think that a single research project might go from a hypothesis to an actual health effect. While this is understandable, given the idealised model, it is not the intention - behind the interactions in the model lie several iterations, and health effects

are usually the results of several research projects. It is also clear that all three impact types are mutually dependant, and that the analysis of one dimension can only provide a partial image of the impact and quality of a given research units production.

How do bibliometric methods apply to the quantifiable aspects of research quality? While this question focuses on bibliometric methods, some other elements will be included as well, in order to portray a more complete concept of research quality. As we know by now, three main dimensions of research quality have been identified: dissemination, policy effects and health effects, all three also considered as types of impact. Other impact types exist too, such as societal impact (Kuruvilla et al., 2006), but these were not discussed by any of the participants in the study, and can likely be classified as other types of impact, that are not directly related to the quality of research. The identified dimensions can be quantitated, and a number of suggestions have been set forth, based on the existing literature as well as results from the interviews and survey. For dissemination, the traditional citation impact is seen as one type of relevant quantitation, although it is still unclear exactly how the proportions of communication and integrated qualities are distributed. As was suggested in section 8.1.13 reference context analysis might provide additional information on this topic. Also the diffusion between basic and applied research, as well as citation paths between different research and study types are potential areas for more intricate analyses of research quality, providing insights in supplement to academic citation impact. Finally, there appears to be a need for journal-based assessment, as a high degree of trust and importance is ascribed to high-esteem journals. The publication channel thus plays an important role in medical research, and despite many sensible, valid and sound bibliometric arguments against assessing articles based on the journals they are published in, it may be a necessity for the scholarly infrastructure to function. The policy effects have been discussed as a stepping stone on the way from dissemination towards health effects, and clinical practice guidelines have been suggested as a means of measuring the implementation of research into practice. Other indicators of such could be systematic reviews and meta-analyses, as briefly discussed in section 4.2. These document types are placed at the very highest levels of evidence, as summaries and syntheses of the best available evidence, as such they are also incredibly useful for clinical practitioners (which was also suggested by a few interviewees). While they may not necessarily be as direct signals of research implementation or policies, they are certainly important steps towards policy formation. As with the clinical guidelines, however; further research is also required in order to document the usefulness of these document types as indicators of policy effects for evaluative bibliometric purposes. The final impact type, the health effects, are not necessarily bibliometrically measurable (although citation path studies of phase III and IV clinical trials have been mentioned as potential health effect indicators), however; a great number of other indicators used especially in clinical epidemiology might provide useful information in coupling with citation path or impact studies. Examples are e.g. the *prevalence* of disease, i.e. the proportion of a population affected by a disease; the

Figure 9.2: Model of impact types and quantitated elements.



incidence rate, i.e. the rate of new disease occurrences; the *mortality* rate, i.e. the rate of patients which die from a disease, the quality of life and the state of general, public health (which could be seen as a combination of some of the other indicators, e.g. quality of life, prevalence of smoking and obesity etc.). As suggested previously, collaborations between clinical epidemiologists and bibliometricians have great potential for revealing information about which studies lead to actual, measurable health effects.

Gathering all of these quantifiable elements of the three impact types, we can construct a model resembling a nomological network (Cronbach & Meehl, 1955), see fig. 9.2, illustrating the relationship between the theoretical constructs, i.e. the impact types, and their quantifiable elements. When we use this vague “resemblance” of a nomological network it is because the model does not conform strictly to all requirements for a true nomological networks, but yet with strong similarities. This particular model is derived from the conceptual model presented in the previous chapter (fig. 8.14), and while they are both concerned with the research quality concept, each model adds individual perspectives. The former model establishes the role of research quality in the scientific system, and the context of the impact relationships. The present model simplifies these relationships and focuses on the quantitation of the conceptual constructs, and the logical relationships between concepts, qualities and quantitation, as proposed by Bunge (1999) (see section 3.1).

Each of the three constructs in this nomological-network-inspired model are dimensions, or qualities, of the main concept, research quality. These qualities are then further quantitated on the next level, which could be further detailed into variables and units for measurement, however; that is not the purpose here, and as the vast array of variables and metrics would obscure the focus, they are therefore not included in the

model. But logically, based on Bunge (1999) and Geisler (2000), the measurements follow from the quantifications present in the model (see also section 3.1), and it should be possible to develop, adapt or apply the entire measurement range to the quantitated elements. We have thus presented a conceptual framework for a more theory-based measurement approach.

9.2. Implications of results

As the scientific society changes, and bibliometric indicators gradually have become an integrated part of large research communities, citations have also changed from being a question of culture to being a part of the academic infrastructure, including bibliometric indicators based on citations, without which the current scientific system could not exist (Wouters, in press), and which places a much greater responsibility on the developers of bibliometric indicators (Wouters, 2013), as they provide not only an infrastructure for scholarly communication, but also for accountability, which ultimately can make and break careers. The results of this dissertation fall in line with these observations, as the articulation of the different dimensions of research quality call for more detailed and diverse assessments of research than pure citation impact analyses. While not every bibliometric research assessment necessarily needs to incorporate several dimensions of research quality, or multiple assessment methods, we should however strive to use more precise terms to describe which aspects of impact or quality are assessed. The result of this would be increased comparability between results that actually describe the same aspect of e.g. different institutions, and a better understanding of how different types of results can supplement each other. We should also strive to incorporate results of other indicator types, e.g. in meta-analytical and synthesis studies. In addition, it is also crucial to emphasise previous statements (van Raan & van Leeuwen, 2002) on the importance of collaborating with domain-experts for interpretation of bibliometric studies.

The distinction between different perceptions of academic impact is also stressed by (Waltman, van Eck, & Wouters, 2013), and especially the conceptual difference between citation impact and a more general, overall “real” impact is discussed here. The results of this dissertation agree well with the theoretical model proposed by Waltman, van Eck and Wouters, but also suffers some of the same difficulties; namely what the “real” impact is, and how we can measure it. In their article a theoretical parameter, α , is defined as the correlation between impact and citedness and ranges from 0 (perfect correlation) to 1 (no correlation). Their model allows a weighted analysis of what “highly cited publications” are, by taking the impact into account, and thus allowing comparisons of researchers’ highly cited publications regardless of the pure citation counts. However, as α is clearly theoretical, there is currently no practical application of the model. In this dissertation we propose some areas for further research, which could potentially give us an indication of the “real” impact of medical research, given a specific definition hereof. These areas are the abovementioned approaches to assess-

ments of health and policy effects, as well as more detailed citation impact studies. At this point the implications would be pure speculation, which is why relatively little energy and space are invested in them. However, these types of detailed, multi-faceted analyses of different impact types might be the path towards better and more responsible citation impact analysis. Here, the focus on these detailed citation impact studies has been on the idea of citation paths, but also the context of citations should be accentuated more clearly. Citation contexts have been used for semi-automatic indexing purposes (Schneider, 2004; Schneider & Borlund, 2005; Schneider, 2006), as signals of paradigm changes (*uniformity*) (Small, 1980) and as symbols of concepts (Small, 1978). Both in the shape of concept symbols and uniformity, citation contexts have been used to signal different uses of references, and through semi-automatic indexing of citation contexts it may be possible to discern between perfunctory use of references and what has been referred to as organic references (Moravcsik & Murugesan, 1975; Small, 1986, 2004; Schneider, 2006), thereby gaining a better impression of which references are used, because they have had a profound impact on the authors of a citing paper, rather than the ceremonial use of perfunctory references. This is interesting from a referencing as well as a citation perspective, but in the research assessment context, it is naturally more obvious to use citation contexts to discern the nature of citations given to assessed articles. Thereby making it possible to calculate a ratio of perfunctory citations for any publication, resulting in a far more exact interpretation of the citation impact of articles. The greatest problem in creating such a ratio is the lack of data availability, as only a rather limited portion of full-text articles are accessible in a format useful for large-scale citation context analysis. New online journals, such as the PLoS journals³⁵ and BioMed Central³⁶, show promising developments in this direction, however, by making machine-readable documents available. Also PubMed Central³⁷ is starting to provide articles in the PubReader format, also allowing XML-downloads of article full-texts (provided that publishers allow this). Further developments in this direction show great promise for the development of bibliometric methods.

An additional perspective on providing more complete impact measurements is provided by the relatively new altmetrics field. Part of the goal of this field is to provide a more diverse evaluation of the impact of research, than citation impact studies can offer (Priem et al., 2010), achieved through measurements of usage of, and social media communication around, research activities. While there are issues for the validity, reproducibility and interpretation of altmetrics (e.g. Wouters & Costas, 2012; Haustein et al., 2013), there is also potential for complementary impact assessment (Priem et al., 2012). Whether altmetrics should be used in parallel with bibliometrics greatly depends on the desired impact assessment, as some types of altmetrics, e.g. usage metrics, might provide good complements to academic citation impact, while e.g. tweets or facebook shares are more likely to reflect a completely different type of impact that is

³⁵<http://www.plos.org/>

³⁶<http://www.biomedcentral.com/>

³⁷<http://www.ncbi.nlm.nih.gov/pmc/>

not necessarily related to any type of academic impact.

With regard to the highest-level research assessment, the quality assessment, it has been clearly established here that citation impact only provides a part of the image. We have also seen that journal esteem plays a role, but only as a part of the communication component of the dissemination dimension; Several other factors play a role in the combined quality of research publications. This partiality stands in sharp contrast to the role journals play in national assessment exercises, such as that of Norway, Denmark and the regional assessment in Flandern, and in the future UK REF, as also discussed in section 2.2.2. To elaborate further on this problem: in Denmark a national system has been implemented, where a quarter of the government funding of universities is distributed according to what is labelled as “quality”³⁸. The difference between this system, and its predecessor is the implementation of a bibliometric indicator, which is thus apparently interpreted as a “quality” indicator, in accordance with the new name of the system. The bibliometric indicator basically assigns scores to articles, according to which journal they are published in³⁹ (FIVU, 2013), and distributes these scores among the affiliated institutions of the authors. The scores are decided by peer panels, and divided into two groups, awarding one and three points respectively. Citation-based journal impact indicators do not officially play a role in the classification of journals, but some indicators are provided to the peer panels as inspiration (FIVU, 2012). Thus universities are awarded funding based on the number of articles they publish, which journals they publish their articles in and how they collaborate. The first and third elements of this system are not related to quality, but quantity and distribution, which means the publishing journal becomes a direct proxy of research quality. This exclusive focus on journals, or publishing channels in general, conflicts not only with the findings of this dissertation, but also with a large body of bibliometric research (see e.g. section 3.2), and in line with the introductory paragraph of this section, bibliometricians need to responsibly challenge such practices. While this is easier said than done, the results of this dissertation illustrate that given the data, it is possible to create far more detailed, large-scale bibliometric assessments, and that we should strive to strategically collaborate on such projects, with the aim of more diverse, reliable indicators of complementary impact indicators. These impact indicators should not merely rely on bibliometric data and methods, but also include domain experts, peer assessments and possibly scientometric indicators of other research aspects. While this will necessarily create a slower model - especially as it can not solely operate on an “expected impact” based on perceived journal esteem - it should also result in a more responsible, valid and sound model for the assessment of research quality.

³⁸In Danish: “Basismidler efter kvalitet”, translated: “Basic funding according to quality”

³⁹There are also mechanisms for scoring books, dissertations, anthologies etc.

10. Final summary

The aim of this dissertation was to conceptualise research quality, thereby articulating the connection between this complex, multi-dimensional concept and evaluative bibliometrics. Medicine was chosen as a domain-delineation of the topic, in order to focus the findings and gain a more intricate understanding of what research quality is. The purpose of this goal was to articulate different dimensions of research quality, in order to provide a better understanding of how it can be quantitated, and how bibliometric measurements of research publications add to these quantitations. The theoretical framework for conceptualisation and quantitation was inspired by Bunge (1999), and the relationship to measurement by Geisler (2000). This was illustrated as logically dependant steps:

concept → *qualities* → *quantitation* → *variables* → *units* → *metrics*

, where the present dissertation focused on conceptualisation (*concept*, *qualities* (dimensions) and *quantitation*), while leaving the measurement part open for further research.

In order to reach this goal, three research questions were posed, and the responses were summarised in the previous chapter. To very briefly recapture, we collected a large number of statements on research quality from interviews with medical researchers, and compiled them into 32 main variables concerned with different aspects of research quality. Using factor analysis, it was possible to identify underlying relationships of variables as well as rank their importance. These factors described different aspects of research quality, but it was through a model of their interactions it was possible to more clearly describe how they each contributed to different dimensions of research quality. The dimensions were labelled dissemination, policy effects and health effects, each of which is also a complex construct in itself. The dissemination dimensions clearly interfaces with bibliometric research methodology, but citation analysis may potentially illustrate elements of the other dimensions as well. However, it is stressed repeatedly that any of these dimensions should not be assessed by bibliometrics alone, and preferably interpretation should be made in collaboration with domain experts. The final aim of conceptualisation was reached through a synthesis of the responses to the research questions,

which resulted in two different models, showing their relationship to the research system and some of their key quantitated elements respectively. These quantitations have implications for the interpretation of existing bibliometric methodology as well as the further development of new or modified methods. With regard to the existing methodology, the results articulate the role of citations in academic impact assessment and call for a stringent terminology for different impact types. Citation impact is also elaborated as a two-component construct, where both communication and internal qualities of research articles play a role in the successful dissemination of research. With regard to further developments, we suggest an extension, or re-interpretation, of Smalls citation pathways (Small, 2004) in combination with Lewison's call for the use of non-journal-article publications as supplementary bibliometric sources (Lewison, 2002a, 2004). The aim of this development is to track policy and health effects of research papers using citation paths, thereby providing additional insight into the complete impact of research. Also the diffusion between basic and applied medical research (biomedical and clinical) appears to be an area worth investigating further, as it could potentially reveal different practices of research groups, and subsequently further inform e.g. funding committees.

We have thus shown a conceptual definition of research quality which incorporates multiple dimensions of impact, which can be quantitated, and their relation to the research process, infrastructure and communication system. In this process we have articulated the need for a more elucidated and stringent impact terminology, and we have outlined the need for systematic studies of diverse impact for a more comprehensive assessment of research quality, not limited to bibliometric evaluations. Finally we have pointed toward the potential of further research in citation diffusion, path and context studies. These developments can help us inform the consumers of bibliometrics better, which in turn should allow them to make more well-informed decisions on e.g. research policy, management and funding.

References

- Abramo, G., & D'Angelo, C. A. (2011). Evaluating research: from informed peer review to bibliometrics. *Scientometrics*, 87(3), 499-514.
- Abramo, G., D'Angelo, C. A., & Di Costa, F. (2011). National research assessment exercises: a comparison of peer review and bibliometrics rankings. *Scientometrics*, 89(3), 929-941.
- Aksnes, D. W. (2003). A macro study of self-citation. *Scientometrics*, 56(2), 235-246.
- Aksnes, D. W. (2006). Citation rates and perceptions of scientific contribution. *Journal of the American Society for Information Science and Technology*, 57(2), 169-185.
- Aksnes, D. W., & Rip, A. (2009). Researchers' perceptions of citations. *Research policy*, 38(6), 895-905.
- Aksnes, D. W., Rorstad, K., Piro, F., & Sivertsen, G. (2011). Are female researchers less cited? A large-scale study of Norwegian scientists. *Journal of the American Society for Information Science and Technology*, 62(4), 628-636.
- Aksnes, D. W., Schneider, J. W., & Gunnarsson, M. (2012). Ranking national research systems by citation indicators. A comparative analysis using whole and fractionalised counting methods. *Journal of informetrics*, 6(1), 36-43.
- Allen, L., Jones, C., Dolby, K., Lynn, D., & Walport, M. (2009). Looking for landmarks: the role of expert review and bibliometric analysis in evaluating scientific publication outputs. *PLoS one*, 4(6), e5910.
- Amr, S., & Tbakhi, A. (2007). Abu Bakr Muhammad Ibn Zakariya Al Razi (Rhazes): Philosopher, physician and alchemist. *Annals of Saudi medicine*, 27(4), 305-307.
- Andersen, J. P. (2013). Association between quality of clinical practice guidelines and citations given to their references. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. F. Moed (Eds.), *Proceedings of the 14th international society of scientometrics and informetrics conference* (Vol. I, p. 215-228). Vienna: Austrian Institute of Technology.
- Andersen, J. P., Christensen, A. L., & Schneider, J. W. (2012). An approach for empirical validation of citation-based journal indicators. In *Proceedings of STI 2012 Montréal*. Montréal, Canada: 17th International Conference on Science and Technology Indicators.
- Andersen, J. P., & Schneider, J. W. (2011). Influence of study design on the citation patterns of Danish, medical research. In E. Noyons, P. Ngulube, & J. Leta (Eds.), *Proceedings of the 13th ISSI Conference* (p. 46-51). Durban, South Africa: 13th ISSI Conference.
- Andras, P. (2011). Research: metrics, quality and management implications. *Research evaluation*, 20(2), 90-106.
- Angell, M. J. (1989). Negative studies. *New England journal of medicine*, 321(7), 464-466.
- Baerlocher, M. O., Newton, M., Gautam, T., Tomlinson, G., & Detsky, A. S. (2007). The meaning of author order in medical research. *Journal of investigative medicine*, 55(4), 174-180.
- Bar-Ilan, J. (2008). Which h-index? - a comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257-271.
- Bates, T., Anić, A., Marusić, M., & Marusić, A. (2004). Authorship criteria and disclosure of contributions: comparison of 3 general medical journals with different author contribution forms. *Journal of the American Medical Association*, 292(1), 86-88.
- Bazeley, P. (2010). Conceptualising research performance. *Studies in higher education*, 35(8), 889-903.
- Bence, V. (2004). The influence of peer review on the Research Assessment Exercise. *Journal of information science*, 30(4), 347-368.
- Bennett, D. M., & Taylor, D. M. (2003). Unethical practices in authorship of scientific papers. *Emergency medicine*, 15(3), 263-270.
- Birukou, A., Wakeling, J. R., Bartolini, C., Casati, F., Marchese, M., Mirylenka, K., ... Wassef, A. (2011). Alternatives to peer review: novel approaches for research evaluation. *Frontiers in computational neuroscience*, 5(56), 12.

- Björk, B. C. (2007). A model of scientific communication as a global distributed information system. *Information research*, 12(2), 307.
- Bollen, J., van de Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS one*, 4(6), e6022.
- Bonzi, S., & Snyder, H. (1991). Motivations for citation: a comparison of self citation and citation to others. *Scientometrics*, 21(2), 245-254.
- Bornmann, L., & Daniel, H. D. (2007). Gatekeepers of science - Effects of external reviewers' attributes on the assessments of fellowship applications. *Journal of informetrics*, 1(1), 83-91.
- Bornmann, L., & Daniel, H. D. (2008a). Selecting manuscripts for a high impact journal through peer review: a citation analysis of Communications that were accepted by Angewandte Chemie International edition, or rejected but published elsewhere. *Journal of the American Society for Information Science & Technology*, 59, 1841-1852.
- Bornmann, L., & Daniel, H. D. (2008b). What do citation counts measure? A review of studies on citing behaviour. *Journal of documentation*, 64(1), 45-80.
- Bornmann, L., & Daniel, H. D. (2010a). The usefulness of peer review for selecting manuscripts for publication: a utility analysis taking as an example a high-impact journal. *PLoS one*, 5(6), e11344.
- Bornmann, L., & Daniel, H. D. (2010b). The validity of staff editors' initial evaluations of manuscripts: a case study of Angewandte Chemie International edition. *Scientometrics*, 85(3), 681-687.
- Bornmann, L., de Moya Anegón, F., & Leydesdorff, L. (2010). Do scientific advancements lean on the shoulders of giants? A bibliometric investigation of the Ortega hypothesis. *PLoS one*, 5(10), e13327.
- Bornmann, L., Mutz, R., & Daniel, H. D. (2008). Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and Technology*, 59(5), 830-837.
- Botts, D. M., & Hargens, L. L. (1991). Are sociologists' publications uncited? Citation rates of journal articles, chapters and books. *The American sociologist*, 63, 147-158.
- Bradford, S. (1985). Sources of information on specific subjects. *Journal of information science*, 10(4), 173-180.
- Brooks, N. (2011). *Demographic characteristics of the House of Delegates and AMA Leadership* (Tech. Rep.). The Council on Long Range Planning and Development, American Medical Association.
- Bryman, A. (1984). The debate about quantitative and qualitative research: A question of method or epistemology? *The British journal of sociology*, 35(1), 75-92.
- Bunge, M. (1993). Realism and antirealism in social science. *Theory and decision*, 35, 207-235.
- Bunge, M. (1999). *The sociology-philosophy connection*. New Brunswick, NJ, USA: Transaction Publishers.
- Bush, V. (1945). As we may think. *The Atlantic Monthly*, 176(July), 101-108.
- Butler, L., & McAllister, I. (2009). Metrics or peer review? Evaluating the 2001 UK Research Assessment Exercise in political science. *Political studies review*, 7(1), 3-17.
- Callahan, M., Wears, R. L., & Weber, E. (2002). Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *Journal of the American Medical Association*, 287(21), 2847-2850.
- Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks : an introduction to co-word analysis. *Social science information*, 22(2), 191-235.
- Campbell, F. M. (1990). National bias: a comparison of citation practices by health professionals. *Bulletin of the Medical Library Association*, 78, 376-382.
- Charlton, B. G. (2010). The cancer of bureaucracy: how it will destroy science, medicine, education; and eventually everything else. *Medical hypotheses*, 74(6), 961-965.
- Clarke, M. (2005). Quality assessment lessons from Australia and New Zealand. *Higher education in Europe*, 30(2), 183-197.

- Clerides, S., Pashardes, P., & Polycarpou, A. (2011). Peer review vs. metric-based assessment: testing for bias in the RAE ratings of UK economics departments. *Economica*, 78(311), 565-583.
- Cole, J. R., & Cole, S. (1971). Measuring the quality of sociological research: Problems in the use of the Science Citation Index. *The American sociologist*, 6, 23-29.
- Cole, J. R., & Cole, S. (1972). The Ortega hypothesis. *Science*, 178(4059), 368-375.
- Cole, S., & Cole, J. R. (1967). Scientific output and recognition : A study in the operation of the reward system of science. *American sociological review*, 32, 377-390.
- Colman, E. (1999). The first English medical journal: Medicina Curiosa. *The Lancet*, 354(9175), 324-326.
- Costas, R., & Bordons, M. (2008). Is g-index better than h-index? An exploratory study at the individual level. *Scientometrics*, 77(2), 267-288.
- Cozzens, S. E. (1981). Taking the measure of science: A review of citation theories. *Newsletter for the International Society for the Sociology of Knowledge*, 7(1&2), 16-21.
- Cozzens, S. E. (1989). What do citations count? The rhetoric-first model. *Scientometrics*, 15(5-6), 437-447.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281-302.
- Cronin, B. (2000). Semiotics and evaluative bibliometrics. *Journal of documentation*, 56(4), 440-453.
- Cronin, B., & Shaw, D. (1999). Citation, funding acknowledgement and author nationality relationships in four information science journals. *Journal of documentation*, 55(4), 402-408.
- Dawes, J. G. (2002). Five point vs. eleven point scale : does it make a difference to data characteristics? *Australasian journal of market research*, 10(1), 39-47.
- Dawes, J. G. (2008). Do data characteristics change according to the number of scale point used? An experiment using 5 point, 7 point and 10 point scales. *International journal of market research*, 50(1), 20.
- De Bellis, N. (2009). *Bibliometrics and citation analysis : from the Science citation index to cybermetrics*. Lanham, Maryland: The Scarecrow Press, Inc.
- Derrick, G. E., Haynes, A., Chapman, S., & Hall, W. D. (2011). The association between four citation metrics and peer rankings of research influence of Australian researchers in six fields of public health. *PLoS one*, 6(4), e18521.
- Easterbrook, P. J., Berlin, J. A., Gopalan, R., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet*, 337(8746), 867-872.
- Egghe, L. (1998). Mathematical theories of citation. *Scientometrics*, 43(1), 57-62.
- Egghe, L., & Rousseau, R. (1990). *Introduction to informetrics. Quantitative methods in library, documentation and information science*. Amsterdam, Netherlands: Elsevier.
- Elton, L. (2000). The UK Research Assessment Exercise: unintended consequences. *Higher education quarterly*, 54(3), 274-283.
- Ernst, E., & Lee, M. S. (2008). A trial design that generates only "positive" results. *Journal of postgraduate medicine*, 54(3), 214-216.
- Estoup, J. (1916). *Gammes sténographique*. Paris: Gauthiers-Villars.
- Etzkowitz, H. (2003). Innovation in innovation: The Triple Helix of University-Industry-Government relations. *Social science information*, 42(3), 293-337.
- Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology*, 4(5), e157.
- Falagas, M. E., & Alexiou, V. G. (2008). The top-ten in journal impact factor manipulation. *Archivum immunologiae et therapeuticae experimentalis*, 56(4), 223-226.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS one*, 4(5), e5738.
- Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891-904.
- Fanelli, D. (2013). Positive results receive more citations, but only in some disciplines. *Scientometrics*, 94(2), 701-709.

- Fearon, J. D., & Laitin, D. D. (2009). Integrating qualitative and quantitative methods. In R. E. Goodin (Ed.), *The oxford handbook of political science* (p. 1166-1186). New York, NY: Oxford University Press, USA.
- Ferket, B. S., Colkesen, E. B., Visser, J. J., Spronk, S., Kraaijenhagen, R. A., Steyerberg, E. W., & Hunink, M. G. M. (2010). Systematic review of guidelines on cardiovascular risk assessment : Which recommendations should clinicians follow for a cardiovascular health check? *Annals of internal medicine*, 170(1), 27-40.
- Ferket, B. S., Genders, T. S. S., Colkesen, E. B., Visser, J. J., Spronk, S., Steyerberg, E. W., & Hunink, M. G. M. (2011). Systematic review of guidelines on imaging of asymptomatic coronary artery disease. *Journal of the American College of Cardiology*, 57(15), 1591-1600.
- Ferket, B. S., Spronk, S., Colkesen, E. B., & Hunink, M. G. M. (2011). Systematic review of guidelines on peripheral artery disease screening. *The American journal of medicine*, 125(15), 198-208.
- FIVU. (2012). *Retningslinjer for niveauinddeling af serier 2012*. Internet. Retrieved from <http://fivu.dk/forskning-og-innovation/statistik-og-analyser/den-bibliometriske-forskningsindikator/niveauinddeling-af-publiceringskanaler/retningslinjer-for-niveauinddeling.pdf>
- FIVU. (2013). *Den bibliometriske forskningsindikator*. Internet. Retrieved from <http://fivu.dk/forskning-og-innovation/statistik-og-analyser/den-bibliometriske-forskningsindikator>
- Flanagin, A., Carey, L. A., Fontanarosa, P. B., Phillips, S. G., Pace, B. P., Lundberg, G. D., & Rennie, D. (1998). Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. *Journal of the American Medical Association*, 280, 222-224.
- Frandsen, T. F. (2004). Journal diffusion factors – a measure of diffusion? *Aslib proceedings*, 56(1), 5-11.
- Frandsen, T. F., Rousseau, R., & Rowland, I. (2006). Diffusion factors. *Journal of documentation*, 62(1), 58-72.
- Freiman, J. A., Chalmers, T. C., Smith Jr, H., & Kuebler, R. R. (1978). The importance of beta, the type ii error and sample size in the design and interpretation of the randomized controlled trial: survey of 71 "negative" trials. *New England journal of medicine*, 299, 690-694.
- Garfield, E. (1955). Citation indexes for science. *Science*, 123(3159), 108-111.
- Garfield, E. (1964). The Citation Index - a new dimension in indexing. *Science*, 144, 649-654.
- Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, 1(4), 359-375.
- Garrison, F. (1966). *History of medicine*. Philadelphia: W.B. Saunders Company.
- Gauffriau, M., Larsen, P. O., Maye, I., Roulin-Periard, A., & von Ins, M. (2007). Publication, cooperation and productivity measures in scientific research. *Scientometrics*, 73(2), 175-214.
- Gauffriau, M., Larsen, P. O., Maye, I., Roulin-Periard, A., & von Ins, M. (2008). Comparisons of results of publication counting using different methods. *Scientometrics*, 77(1), 147-176.
- Geisler, E. (2000). *The metrics of science and technology*. Westport, CT, USA: Quorum Books.
- Gerring, J. (2012). *Social science methodology - a unified framework* (2nd ed.; C. Elman, J. Gerring, & J. Mahoney, Eds.). New York: Cambridge University Press.
- Gilbert, G. (1977). Referencing as persuasion. *Social studies of science*, 7(1), 113-122.
- Glasziou, P., Vandenbroucke, J., & Chalmers, I. (2004). Assessing the quality of research. *British medical journal*, 328, 39-41.
- Glänzel, W. (1996). The need for standards in bibliometric research and technology. *Scientometrics*, 35(2), 167-176.
- Glänzel, W., & Schoepflin, U. (1994). Little Scientometrics, Big Scientometrics ...and beyond? *Scientometrics*, 30(2-3), 375-384.
- Gläser, J., & Laudel, G. (2007). The social construction of bibliometric evaluations. In R. Whitley & J. Gläser (Eds.), *The changing governance of the sciences* (p. 101-123). Dordrecht, Netherlands: Springer+Business Media B.V.

- Goertz, G. (2006). *Social science concepts: A user's guide*. Princeton, New Jersey: Princeton University press.
- Goldstein, H. (2011). Estimating research performance by using research grant award gradings. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(1), 83-93.
- Goodman, S. N. (1999). Toward evidence-based medical statistics. 1: The p value fallacy. *Annals of internal medicine*, 130, 995-1004.
- Goodman, S. N. (2008). A dirty dozen: twelve p-value misconceptions. *Seminars in hematology*, 45(3), 135-140.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Grange, R. I. (1999). National bias in citations in urology journals: parochialism or availability? *BJU international*, 84(6), 601-603.
- Greenhalgh, T. (2010). *How to read a paper: The basics of evidence-based medicine* (4th ed.). Oxford: BMJ Books.
- Greenwood, D. C. (2011). Reliability of journal impact factor rankings. *BMC medical research methodology*, 7, 48.
- Gøtzsche, P. C., Hrobjartsson, A., Johansen, H. K., Haahr, M. T., Altman, D. G., & Chan, A. W. (2007). Ghost authorship in industry-initiated randomised trials. *PLoS medicine*, 4, e19.
- Harremoës, P., & Topsøe, F. (2006). Zipf's Law, hyperbolic distributions and entropy loss. *Lecture notes in computer science*, 4123, 788-792.
- Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2013). Coverage and adoption of altmetrics sources in the bibliometric community. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. F. Moed (Eds.), *Proceedings of the 14th international society of scientometrics and informetrics conference* (Vol. I, p. 468-483). Vienna: Austrian Institute of Technology.
- Hjørland, B. (2002). Domain analysis in information science: Eleven approaches - traditional as well as innovative. *Journal of documentation*, 58(4), 422-462.
- Huang, M. H., Lin, C. S., & Chen, D. Z. (2011). Counting methods, country rank changes, and counting inflation in the assessment of national research productivity and impact. *Journal of the American Society for Information Science and Technology*, 62(12), 2427-2436.
- Hudson Jones, A. (2003). Can authorship policies help prevent scientific misconduct? What role for scientific societies? *Science and engineering ethics*, 9(2), 243-256.
- Hunter, D. (2008). Health needs more than health care: the need for a new paradigm. *European journal of public health*, 18(3), 217-219.
- Ibáñez, A., Larrañaga, P., & Bielza, C. (2011). Using Bayesian networks to discover relationships between bibliometric indices. A case study of computer science and artificial intelligence journals. *Scientometrics*, 89(2), 523-551.
- ICMJE. (1988). Uniform requirements for manuscripts submitted to biomedical journals. *British Medical Journal*, 296, 401-405.
- Informa Healthcare. (2005). *ICH Harmonised Tripartite Guideline for GCP*. London: Informa.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.
- Joyce, J. (1922). *Ulysses*. Paris: Sylvie Beach.
- Kahn, R., & Kahn, P. (1997). The Medical Repository - the first U.S. medical journal (1797-1824). *The New England journal of medicine*, 337(26), 1926-1930.
- Kaplan, N. (1965). The norms of citation behavior: Prologomena to the footnote. *American documentation*, 16(3), 179-184.
- Kassirer, J. P., & Champion, E. W. (1994). Peer review : crude and understudied, but indispensable. *The journal of the American Medical Association*, 272(2), 96-97.
- Katz, J. S. (1999). The self-similar science system. *Research policy*, 28(5), 501-517.
- Kemper, T. D. (1981). Social constructionist and positivist approaches to the sociology of emotions. *American journal of sociology*, 87(2), 336-362.

- Kenna, R., & Berche, B. (2011). Normalization of peer-evaluation measures of group research quality across academic disciplines. *Research evaluation*, 20(2), 107-116.
- Kiss, G. R. (1975). An associative Thesaurus of English: structural analysis of a large relevance network. In A. Kennedy & A. Wileks (Eds.), *Studies in long term memory* (p. 103-121). Wiley.
- Kjaergaard, L. L., & Gluud, C. (2002). Citation bias of hepato-biliary randomized clinical trials. *Journal of clinical epidemiology*, 55(4), 407-410.
- Kostoff, R. N., & Geisler, E. (2007). The unintended consequences of metrics in technology evaluation. *Journal of informetrics*, 1(2), 103-114.
- Kotrc, R., & Walters, K. (1979). A bibliography of the Galenic Corpus. A newly researched list and arrangement of the titles of the treatises extant in Greek, Latin, and Arabic. *Transactions & studies of the College of Physicians of Philadelphia*, 1(4), 256-304.
- Kuruvilla, S., Mays, N., Pleasant, A., & Walt, G. (2006). Describing the impact of health research: a research impact framework. *BMC health services research*, 6, 134.
- Kvale, S. (2009). *Interview*. Copenhagen: Gyldendal Akademisk.
- Landes, W. M., & Posner, R. A. (2000). Citations, age, fame, and the web. *The journal of legal studies*, 29(S1), 319-344.
- Larivière, V., & Gingras, Y. (2009). The impact factor's Matthew Effect: a natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology*, 61(2), 424-427.
- Latour, B. (1987). *Science in action : How to follow scientists and engineers through society*. Milton Keynes: Open University Press.
- Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1), 2-17.
- Lehmann, S., Jackson, A. D., & Lautrup, B. (2006). Measures for measures. *Nature*, 444, 1003-1004.
- Lehmann, S., Jackson, A. D., & Lautrup, B. (2008). A quantitative analysis of indicators of scientific performance. *Scientometrics*, 76(2), 369-390.
- LERU. (2012). *Research universities and research assessment* (Tech. Rep.). Leuven, Belgium: League of European Research Universities.
- Lewison, G. (2002a). From biomedical research to health improvement. *Scientometrics*, 54(2), 179-192.
- Lewison, G. (2002b). Researchers' and users' perceptions of the relative standing of biomedical papers in different journals. *Scientometrics*, 53(2), 229-240.
- Lewison, G. (2003). Beyond outputs: new measures of biomedical research impact. *Aslib proceedings*, 55(1/2), 32-42.
- Lewison, G. (2004). Citations to papers from other documents. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research: the use of publication and patent statistics in studies of s&T systems* (p. 457-472). Dordrecht: Kluwer Academic Publishers.
- Lewison, G., & Dawson, G. (1998). The effect of funding on the outputs of biomedical research. *Scientometrics*, 41(1-2), 17-27.
- Leydesdorff, L. (1987). Towards a theory of citation? *Scientometrics*, 12(5-6), 305-309.
- Leydesdorff, L. (1989). Words and co-words as indicators of intellectual organization. *Research policy*, 18(4), 209-223.
- Leydesdorff, L. (1997). Why words and co-words cannot map the development of the sciences. *Journal of the American Society for Information Science*, 48(5), 418-427.
- Leydesdorff, L. (1998). Theories of citation? *Scientometrics*, 43(1), 5-25.
- Leydesdorff, L. (2009). How are new citation-based journal indicators adding to the bibliometric toolbox? *Journal of the American Society for Information Science & Technology*, 60(7), 1327-1336.
- Leydesdorff, L., & Bornmann, L. (2011). How fractional counting of citations affects the impact factor: Normalization in terms of differences in citation potentials among fields of science. *Journal of the American Society for Information Science and Technology*, 62(2), 217-229.

- Leydesdorff, L., Bornmann, L., Mutz, R., & Opthof, T. (2011). Turning the tables on citation analysis one more time. *Journal of the American Society for Information Science and Technology*, 62(7), 1370-1381.
- Leydesdorff, L., & Etzkowitz, H. (1996). Emergence of a Triple Helix of university-industry-government relations. *Science and public policy*, 23(5), 279-286.
- Leydesdorff, L., & Opthof, T. (2010a). Normalization at the field level: fractional counting of citations. *Journal of informetrics*, 4(4), 644-646.
- Leydesdorff, L., & Opthof, T. (2010b). Scopus' Source Normalized Impact per Paper (SNIP) versus the Journal Impact Factor based on fractional counting of citations. *Journal of the American Society for Information Science and Technology*, 61(11), 2365-2396.
- Lokker, C., McKibbin, K. A., McKinlay, R. J., Wilczynski, N. L., & Haynes, R. B. (2008). Prediction of citation counts for clinical articles at two years using data available within three weeks of publication: retrospective cohort study. *British medical journal*, 336(7645), 655-657.
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12), 317-323.
- Lundberg, G. A. (1939). Contemporary positivism in sociology. *American sociological review*, 4(1), 42-55.
- Luukkonen, T. (1997). Why has Latour's theory of citation been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis. *Scientometrics*, 38, 27-37.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological methods*, 4(1), 84-99.
- MacRoberts, M. H., & MacRoberts, B. R. (1987). Testing the Ortega hypothesis: facts and artifacts. *Scientometrics*, 12(5-6), 293-295.
- MacRoberts, M. H., & MacRoberts, B. R. (1989). Another test of the normative theory of citing. *Journal of the American Society for Information Science*, 16(4), 151-172.
- MacRoberts, M. H., & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics*, 36(3), 435-444.
- Marincola, F. (2003). Translational Medicine: A two-way road. *Journal of translational medicine*, 1(1), 1.
- Markel, H. (2004). "i swear by Apollo" - on taking the Hippocratic oath. *The New England journal of medicine*, 350(20), 2026-2029.
- McKibbin, K. A., Wilczynski, N. L., & Haynes, R. (2004). What do evidence-based secondary journals tell us about the publication of clinically important articles in primary healthcare journals? *BMC medicine*, 2, 33.
- McMullin, E. (1984). A case for scientific realism. In J. Leplin (Ed.), *Scientific realism* (p. 8-40). Berkeley, CA: University of California Press.
- Merton, R. K. (1942). *The normative structure of science*. reprinted in Merton, The Sociology of Science, Norman W. Storer, ed. (Chicago: University of Chicago Press, 1973).
- Merton, R. K. (1968). The Matthew Effect in science. *Science*, 159(3810), 56-63.
- Merton, R. K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, IL: University of Chicago Press.
- Meyeroth, M. (1935). Thirty-three clinical observations by Rhazes (circa 900AD). *Isis*, 23(2), 321-372.
- Miles, A. (2009). On a medicine of the whole person: away from scientific reductionism and towards the embrace of the complex in clinical practice. *Journal of evaluation in clinical practice*, 15(6), 941-949.
- Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht, The Netherlands: Springer.
- Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of informetrics*, 4(3), 265-277.

- Moed, H. F., & van Leeuwen, T. (1995). Improving the accuracy of Institute for Scientific Information's Journal Impact Factors. *Journal of the American Society for Information Science*, 46(6), 461-467.
- Moed, H. F., & van Leeuwen, T. (1996). Impact factors can mislead. *Nature*, 381, 186.
- Moerner, W. E., & Orrit, M. (1999). Illuminating single molecules in condensed matter. *Science*, 283(5408), 1670-1676.
- Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social studies of science*, 5, 86-92.
- Morris, A. H., & Ioannidis, J. P. A. (2013). Limitations of medical research and evidence at the patient-clinician encounter scale. *Chest*, 143(4), 1127-1135.
- Mory, R. N., Mindell, D., & Bloom, D. A. (2000). The Leech and the Physician: Biology, etymology, and medical practice with Hirudinea medicinalis. *World journal of surgery*, 24, 878-883.
- Moskowitz, D., & Bodenheimer, T. (2011). Moving from evidence-based medicine to evidence-based health. *Journal of general internal medicine*, 26(6), 658-660.
- Nature neuroscience editors. (2005). Revolutionizing peer review? [editorial]. *Nature neuroscience*, 8, 397.
- Navarro Puerto, M. A., Ibarluzea, I. G., Ruiz, O. G., Alvarez, F. M., Herreros, R. G., Pintiado, R. E., ... León, I. M. (2008). Analysis of the quality of clinical practice guidelines on established ischemic stroke. *International journal of technology assessment in health care*, 24(3), 333-341.
- Nielsen, J. (1992). Finding usability problems through heuristic evaluation. In *Proceedings of the sigchi conference on human factors in computing systems - chi '92* (p. 373-380). New York: ACM Press.
- Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the sigchi conference on human factors in computing systems empowering people - chi '90* (p. 249-256). New York: ACM Press.
- Nielsen, M. L., & Ingwersen, P. (1999). The word association methodology - a gateway to work-task based retrieval. In S. W. Draper, M. D. Dunlop, I. Ruthven, & C. J. Van Rijsbergen (Eds.), *Proceedings of mira 99: Evaluating interactive information retrieval* (p. 17-27). Glasgow.
- Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education*, 15(5), 625-632.
- Norris, M., & Oppenheim, C. (2011). Peer review and the h-index: Two studies. *Journal of informetrics*, 4(3), 221-232.
- Nutton, V. (2006). The rise of medicine. In R. Porter (Ed.), *The cambridge history of medicine* (p. 46-70). New York: Cambridge University Press.
- OCEBM Levels of Evidence Working Group. (2011). *The oxford 2011 levels of evidence*. Oxford: Oxford Centre for Evidence-Based Medicine.
- Opatrný, T. (2008). Playing the system to give low-impact journals more clout. *Nature*, 455, 50676.
- Ophof, T., & Leydesdorff, L. (2010). Caveats for the journal and field normalizations in the CWTS (Leiden) evaluations of research performance. *Journal of informetrics*, 4(3), 423-430.
- Overbeke, J. (1999). The state of evidence: what we know and what we don't know about journal peer review. In F. Godlee & T. Jefferson (Eds.), *Peer review in health sciences* (1st ed., p. 32-44). London: BMJ Books.
- Parinov, S., & Neylon, C. (2011). Science as a social system and virtual research environment. *Journal of artificial societies and social stimulation*, 14(4), 10.
- Patsopoulos, N. A., Analatos, A. A., & Ioannidis, J. P. A. (2005). Relative citation impact of various study designs in the health sciences. *JAMA: the journal of the American Medical Association*, 293(19), 2362-2366.
- Perneger, T. V. (2010). Citation analysis of identical consensus statements revealed journal-related bias. *Journal of clinical epidemiology*, 63(6), 660-664.
- Piowar, H. A., & Priem, J. (2013). The power of altmetrics on a cv. *Bulletin of the Association for Information Science and Technology*, 39(4), 10-13.

- Pons, J., Sais, C., Illa, C., Méndez, R., Suñen, E., Casas, M., & Camí, J. (2010). Is there an association between the quality of hospitals' research and their quality of care? *Journal of health services research & policy*, 15(4), 204-209.
- Price, D. J. S. (1963). *Little Science, Big Science*. New York: Columbia University Press.
- Price, D. J. S. (1965). Networks of scientific papers. *Science*, 149(3683), 510-515.
- Price, D. J. S. (1970). Citation measures of hard science, soft science, technology and non-science. In C. E. Nelson & D. K. Pollock (Eds.), (p. 3-22). Lexington, Mass.: D. C. Heath & Co.
- Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the Wild: using social media to explore scholarly impact. *arXiv*, 1203.4745.
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). *Altmetrics: a manifesto*. online. Retrieved from <http://altmetrics.org/manifesto>
- Prioreschi, P. (2002). Al-Kindi, a precursor of the scientific revolution. *Journal of the International Society for the History of Islamic Medicine*, 1(2), 17-20.
- Pudovkin, A. I., & Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. *Journal of the American Society for Information Science and Technology*, 53(13), 1113-1119.
- R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org> (ISBN 3-900051-07-0)
- Raiche, G. (2010). nFactors: an R package for parallel analysis and non graphical solutions to the Cattell scree test [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=nFactors> (R package version 2.3.3.)
- Raiche, G., Walls, T. A., Magis, D., Riopel, M., & Blais, J. G. (2013). Non-graphical solutions for Cattell's scree test. *Methodology: European journal of research methods for the behavioral and social sciences*, 9(1), 23-29.
- Ragin, C. C. (2000). *Fuzzy-set social science*. Chicago, ILL: University of Chicago Press.
- Ragin, C. C. (2008). *Redesigning social inquiry : fuzzy sets and beyond*. Chicago, ILL: University of Chicago Press.
- Rasmussen, P. G., & Andersen, J. P. (2013). Altmetrics : an alternate perspective on research evaluation. *ScieComInfo*, 9(2), 1.
- Reich, D. A., Green, M. C., Brock, T. C., & Tetlock, P. E. (2007). Biases in research evaluation: Inflated assessment, oversight, or error-type weighting? *Journal of experimental social psychology*, 43(4), 633-640.
- Reidenberg, M. M. (1989). Padding the curriculum vitae : Impairing the utility and integrity of the biomedical literature. *Clinical pharmacology & therapeutics*, 45(4), 337-339.
- Rennie, D. (1999). Editorial peer review: its development and rationale. In F. Godlee & T. Jefferson (Eds.), *Peer review in health sciences* (1st ed., p. 3-13). London: BMJ Books.
- Rip, A. (1997). Qualitative conditions of scientometrics: the new challenges. *Scientometrics*, 38(1), 7-26.
- Rolfe, G. (2006). Validity, trustworthiness and rigour: quality and the idea of qualitative research. *Journal of advanced nursing*, 53(3), 304-310.
- Sackett, D. L., Rosenberg, W. C., & Gray, J. A. M. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal*, 312, 71-72.
- Salton, G. (1963). Associative document retrieval techniques using bibliographic information. *Journal of the Association for Computing Machinery*, 10, 440-457.
- Salton, G. (1971). *The SMART Retrieval system*. Englewood Cliffs, NJ, USA: Prentice-Hall.
- Schneider, J. W. (2004). *Verification of bibliometric methods' applicability for thesaurus construction*. Unpublished doctoral dissertation, Royal School of Library and Information Science, Aalborg.
- Schneider, J. W. (2006). Concept symbols revisited : Naming clusters by parsing and filtering of noun phrases from citation contexts of concept symbols. *Scientometrics*, 68(3), 573-593.
- Schneider, J. W. (2013). Caveats for using statistical significance tests in research assessments. *Journal of informetrics*, 7(1), 50-62.

- Schneider, J. W., & Borlund, P. (2005). A bibliometric-based semi-automatic approach to identification of candidate thesaurus terms: Parsing and filtering of noun phrases from citation contexts. *Lecture notes in computer science*, 3507, 226-237.
- Schneider, J. W., & Borlund, P. (2007). Matrix comparison, Part 1: Motivation and important issues for measuring the resemblance between proximity measures or ordination results. *Journal of the American Society for Information Science and Technology*, 58(11), 1586-1595.
- Schutte, H., & Svec, J. (2007). Reaction of Folia Phoniatica et Logopaedica on the current trend of impact factor measures. *Folia phoniatica et logopaedica*, 59(6), 281-285.
- Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, 43(9), 628-638.
- Seglen, P. O. (1997). Why the impact factor should not be used for evaluating research. *British Medical Journal*, 314(7079), 498-502.
- Shaw, G. M., & Chase, J. G. (2007). Why evidence based medicine may be bad for you and your patients. In V. Nayyar (Ed.), *Critical care update* (p. 9-20). New Delhi: Jaypee Medical Publishers.
- Small, H. (1978). Cited documents as concept symbols. *Social studies of science*, 8(3), 327-340.
- Small, H. (1980). Co-citation context analysis and the structure of paradigms. *Journal of documentation*, 36(3), 183-196.
- Small, H. (1986). The synthesis of specialty narratives from co-citation clusters. *Journal of the American Society for Information Science*, 37(3), 97-110.
- Small, H. (2000). Web of knowledge a festschrift in honor of eugene garfield. In (p. 449-473). Information Today Inc.
- Small, H. (2004). On the shoulders of Robert Merton: towards a normative theory of citation. *Scientometrics*, 60(1), 71-79.
- Smart, C. J. (2005). Attributes of exemplary research manuscripts employing quantitative analyses. *Research in higher education*, 46, 461-477.
- Smith, R. (1988). Problems with peer review and alternatives. *British Medical Journal*, 296(6624), 774-777.
- Smith, R. (1997). Journal accused of manipulating impact factor. *British Medical Journal*, 314, 461.
- Smith, S., Ward, V., & House, A. (2011). "impact" in the proposals for the UK's Research Excellence Framework: Shifting the boundaries of academic autonomy. *Research policy*, 40(10), 1369-1379.
- Sønderstrup-Andersen, E. M., & Sønderstrup-Andersen, H. H. K. (2008). An investigation into diabetes researcher's perceptions of the journal impact factor - reconsidering evaluating research. *Scientometrics*, 76(2), 391-406.
- Spiteri, L. (2002). Word association testing and thesaurus construction: Defining inter-term relationships. In *Proceedings of the 30th annual conference of the canadian association for information science* (p. 24-33). Toronto.
- Spärck-Jones, K. (1971). *Automatic keyword classification for information retrieval*. London, England: Butterworth & Co. (Publishers) Ltd.
- Starbuck, W. H. (2006). *The production of knowledge*. Oxford, UK: Oxford University Press.
- Steen, R. G. (2011). Retractions in the scientific literature: do authors deliberately commit research fraud? *Journal of Medical Ethics*, 37(2), 113-117.
- Stern, R. (1990). Uncitedness in the biomedical literature. *Journal of the American Society for Information Science*, 41(3), 193-196.
- Sternberg, R., & Gordeeva, T. (1996). The anatomy of impact: what makes an article influential? *Psychological science*, 7(2), 69-75.
- Strange, K. (2008). Authorship: why not just toss a coin? *American journal of physiology. Cell physiology*, 295(3), C567-75.
- Suls, J., & Martin, R. (2009). The air we breathe: A critical look at practices and alternatives in the peer-review process. *Perspectives on psychological science*, 4(1), 40-50.

- Syed, I. (2002). Islamic medicine: 1000 years ahead of its times. *Journal of the International Society for the History of Islamic Medicine*, 2, 2-9.
- Taylor, J. (2011). The assessment of research quality in UK universities: peer review or metrics? *British journal of management*, 22(2), 202-217.
- Todorov, R., & Glänzel, W. (1988). Journal citation measures: a concise review. *Journal of information science*, 14(1), 47-56.
- Tripp, O., & Feitelson, D. G. (2007). *Zipf's law revisited* (Tech. Rep. No. 2007-115). School of Computer Science and Engineering, The Hebrew University of Jerusalem.
- van den Berg, T., Engelhardt, E. G., Haanstra, T. M., Langius, A. E., & van Tulder, M. W. (2012). Methodology of clinical nutrition guidelines for adult cancer patients: How good are they according to AGREE criteria? *Journal of parenteral and enteral nutrition*, 36(3), 316-322.
- van Leeuwen, T. (2008). Testing the validity of the Hirsch-index for research assessment purposes. *Research evaluation*, 17(2), 157-160.
- van Leeuwen, T., Costas, R., Calero-Medina, C., & Visser, M. (2012). The role of editorial material in bibliometric research performance assessments. In Árchambault, Y. Gingras, & V. Larivière (Eds.), *Proceedings of sti 2012 montréal* (p. 511-522). Montréal, Canada: 17th International Conference on Science and Technology Indicators.
- van Leeuwen, T., & Moed, H. F. (2005). Characteristics of journal impact factors: The effects of uncitedness and citation distribution on the understanding of journal impact factors. *Scientometrics*, 63(2), 357-371.
- van Raan, A. F. J. (1998a). The influence of international collaboration on the impact of research results. *Scientometrics*, 42(3), 423-428.
- van Raan, A. F. J. (1998b). In matters of quantitative studies of science the fault of theorists is offering too little and asking too much. *Scientometrics*, 43(1), 129-139.
- van Raan, A. F. J. (2004). Sleeping Beauties in science. *Scientometrics*, 59(3), 461-466.
- van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67(3), 491-502.
- van Raan, A. F. J., & van Leeuwen, T. (2002). Assessment of the scientific basis of interdisciplinary, applied research : application of bibliometric methods in nutrition and food research. *Research policy*, 31, 611-632.
- van Raan, A. F. J., van Leeuwen, T., Visser, M. S., van Eck, N. J., & Watlman, L. (2010). Rivals for the crown: Reply to Opthof and Leydesdorff. *Journal of informetrics*, 4(3), 431-435.
- Vinkler, P. (1988). An attempt of surveying and classifying bibliometric indicators for scientometric purposes. *Scientometrics*, 13(5-6), 239-259.
- Waltman, L., & Costas, R. (2013). F1000 recommendations as a new data source for research evaluation: A comparison with citations. *CoRR*, abs/1303.3875. Retrieved from <http://arxiv.org/abs/1303.3875>
- Waltman, L., van Eck, N. J., van Leeuwen, T., & Visser, M. S. (2013). Some modifications to the SNIP journal impact indicator. *Journal of informetrics*, 7, 272-285.
- Waltman, L., van Eck, N. J., & Wouters, P. (2013). Counting publications and citations: Is more always better? *Journal of informetrics*, 7(3), 635-641.
- Weber, K. (2007). Science wars or the need for a non-dogmatic defense of realism. In D. Aleksandrowicz & K. Weber (Eds.), *Kulturwissenschaften im blickfeld der standortbestimmung, legitimierung und selbstkritik* (p. 233-254). Berlin: Frank & Timme.
- Weingart, P. (2011). Impact of bibliometrics upon the science system: inadvertent consequences? *Scientometrics*, 62(1), 117-131.
- Weiss, R. S. (1994). *Learning from strangers: The art and method of qualitative interview studies*. New York: Free Press.
- West, R., & McIlwaine, A. (2002). What do citation counts count for in the field of addiction? An empirical evaluation of citation counts and their link with peer ratings of quality. *Addiction*, 87, 501-504.

Conceptualising research quality in medicine for evaluative bibliometrics

- White, H. D. (2004). Reward, persuasion, and the Sokal Hoax: A study in citation identities. *Scientometrics*, 60(1), 93-120.
- White, H. D. (2007a). Combining bibliometrics, information retrieval, and relevance theory, part 1 : First examples of a synthesis. *Journal of the American Society for Information Science & Technology*, 58, 536-559.
- White, H. D. (2007b). Combining bibliometrics, information retrieval, and relevance theory, part 2: Some implications for information science. *Journal of the American Society for Information Science & Technology*, 58, 583-605.
- Williams, G. (1998). Misleading, unscientific, and unjust: the United Kingdom's research assessment exercise. *British Medical Journal*, 316, 1079-1088.
- Wislar, J. S., Flanagan, A., Fontanarosa, P. B., & DeAngelis, C. D. (2011). Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. *British Medical Journal*, 343, d6128.
- World Medical Association. (2002). Declaration of Helsinki: ethical principles for medical research involving human subjects. *Journal of postgraduate medicine*, 48, 206.
- Wouters, P. F. (1997). Citation cycles and peer review cycles. *Scientometrics*, 38(1), 39-55.
- Wouters, P. F. (1999). *The citation culture*. Unpublished doctoral dissertation, University of Amsterdam.
- Wouters, P. F. (2013). *Vice rector university of vienna calls for a new scientometrics*. Internet. Retrieved from <http://citationculture.wordpress.com/>
- Wouters, P. F. (in press). *The citation from culture to infrastructure*.
- Wouters, P. F., & Costas, R. (2012). Users, narcissism and control - tracking the impact of scholarly publications in the 21st century. In . Archambault, Y. Gingras, & V. Larivière (Eds.), *Proceedings of sti 2012* (p. 846-857). Montréal, Canada: STI.
- Yu, G., Yang, D.-H., & Liang, W. (2009). Reliability-based citation impact factor and the manipulation of impact factor. *Scientometrics*, 83(1), 259-270.
- Zhuang, X., Bartley, L. E., Babcock, H. P., Russell, R., Ha, T., Herschlag, D., & Chu, S. (2000). A single-molecule study of RNA catalysis and folding. *Science*, 288(5473), 2048-2051.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, England: Addison-Wesley.
- Zuccala, A. (2010). The mathematical review system: does reviewer status play a role in the citation process? *Scientometrics*, 84(1), 221-235.
- Zuckerman, H. (1987). Citation analysis and the complex problem of intellectual influence. *Scientometrics*, 12, 329-338.

Appendices

A. Interview data

In the following appendix, raw transcriptions of interviews are presented. The transcriptions were created manually, from recorded, digital sound files. Minor differences between transcriptions and sound files exist in cases where a) the dialogue was purely small-talk or introductory, containing no information regarding the subject at all or b) the actual dialogue to a very great degree would not be comprehensible in written form. In case a) short passages have been left out, while cases b) have been re-worded to better fit written language.

A.1. Pilot study transcriptions

A.1.1. Interview reference: P01

Date: Sep 09, 2011

Time: 10:00

Question: *Jeg har bedt dig tage nogle artikler med, som du vurderede som værende af god forskningskvalitet, høj kvalitet. Så hvis du vil fortælle mig lidt om hvorfor du har valgt dem?*

Response: De er jo valgt ud fra det emne, eller det felt jeg arbejder med indenfor statistik, og den første (1), den er om Bayesian forecasting, og den er skrevet af to personer, der er meget kendt indenfor det her felt med state-space modeller, og det som jeg synes er godt ved den er, at den er meget pædagogisk, der er en rød tråd og at de kommer omkring mange detaljer og har både teori og små eksempler med inde i det. Og ikke nok med det, så slutter den af med en kæmpe diskussion hvor det er alle mulige henvendelser fra forskellige andre statsitikere, der har skrevet til dem hvor der så er uddybet med hvad der er de mener er diskussionspunkter til deres artikel, og der er

en hel masse indslag, og så til sidst der svarer de så samlet set på alle de her henvendelser de har fået. Så man får en rigtig god diskussion af det, hele vejen igennem. Og så er den ekstremt brugbar for mit projekt.

Question: *Du siger brugbar for dit projekt, er det så kun i forhold til dit projekt specifikt eller er det bare sådan en artikel der generelt er meget brugbar hvis man beskæftiger sig med det område?*

Response: Jeg vil sige det sidste, altså at det er en der er meget brugbar når man arbejder indenfor det her. Altså det den her artikel specifikt tager fat i er to forskellige metoder til en form for modelselektion hvor det nærmest er det første der er blevet publiceret indenfor det her. Og på trods af at det er den første publikation indenfor det her emne, har de alligevel formået at gøre det sådan at det er let tilgængeligt og let forståeligt når man ikke har læst om det før, og det synes jeg er meget imponerende.

Question: *Så de grundlægger en teori, eller?*

Response: Eller, de bygger videre på en teori. Altså, man har state-space modellerne, det er sådan meget almindeligt kendt, det har været kendt i mange år, det de så går videre med det er en måde til at finde ud af hvilken model der passer bedst til de data jeg sidder med lige nu. Og den slags modelselektion som de har lavet her har man ikke tidligere haft kendskab til indenfor de applikationer vi bruger den til. Derfor er den god, fordi den er nede på jorden, men får det også hævet op så det er noget man kan bruge. Og den næste artikel (2), det er egentlig det samme, hvor de også arbejder med et meget kendt felt, og så dykker de ned i et specifikt, lille område, hvor opbygningen på artiklen også er meget pædagogisk og kommer rundt og forklarer hvorfor de gør som de gør, hvilket tit og ofte kan være et problem med artikler, i hvert fald indenfor denne statistik-verden; at mange ting bliver de nødt til at sige, at det kan vises, at det holder, men de her viser det rent faktisk. Og lige netop denne her artikel er ekstremt brugbar for mit ph.d. fordi det egentlig er det jeg skal ind og anvende, lige netop det de bruger her. Og det jeg også synes er godt ved begge artikler det er, at de kan give mig en idé om hvordan jeg skal bygge mine egne artikler op, med dispositioner og det hele. Og så det at de forfattere, der er på de her artikler er nogle jævnt store kanoner indenfor det her, og alligevel holder de det på et niveau hvor almindelige ph.d.-studerende kan være med. Så de her to, i hvert fald (1) er grundstenen indenfor det her emne, og (2) er den hvor forfatterne til den har skrevet en bog hvor de introducerer en meget bredere form af det de laver i den her artikel. Og forfatterne på (2) er virkelig hotshots indenfor det her, ligesom forfatterne på (1).

Question: *Det var sådan set det der var om dem. Det var jo interessant at høre, hvad du vurderer som kvalitet i den sammenhæng. Det var så meget specifikt i forhold til de her to artikler, men hvis vi nu skal sige det mere generelt... hvis jeg siger forskningskvalitet som begreb, hvad tænker du så på?*

Response: At det er, for mig at se, så er det gode forskningsartikler hvor du sætter et generelt setup og anvender det specifikt på et eller andet dataeksempel, så at du kan tage den teori der er beskrevet i artiklen og kan anvende den på dit eget, ud fra hvordan de har anvendt det på deres eksempel. Det synes jeg er meget brugbart.

Question: *Hvis man nu skulle prøve at evaluere, nu snakker du jo meget om forskningsartikler, og forskning kunne måske også være noget andet, men lad os nu holde os til artiklerne lige nu i hvert fald, hvis man nu skulle evaluere hvor gode de er, ud fra objektive kriterier, er der noget du forestiller dig man kunne gøre?*

Response: Man kunne læse dem igennem og se om de er gode eller dårlige... det ved jeg ikke.

Question: *Så peer review med andre ord?*

Response: Ja, både og, fordi jeg synes der er en fare i det med peer review, fordi det kan afhænge meget af hvilken peer reviewer du får. Og de skal i hvert fald vælges omhyggeligt, sådan nogle reviewers, og så det med at mange artikler, hvor de har et negativt fund, de bliver egentlig forkastet, det synes jeg er lidt uheldigt. Men altså jo, helt klart, der skal da være en eller anden kontrol med folk der læser det igennem, og der skal man have nogen til at læse artiklerne, både nogen der er teoretisk stærke og nogen der er mere anvendelses-stærke i den forstand at de skal kunne opdage hvis der er fejl i teorien, eller de skal opdage hvis der er sjusk med data eller sådan et eller andet. Så på den måde opnår du bedre kvalitet af artiklerne hvis du har en reiew process, hvis reviewerne tager det seriøst og er godt nok udvalgt.

Question: *Hvis vi nu skulle snakke om forskning helt generelt. Lad os nu sige, at vi har to forskellige universiteter, og vi vil gerne vide noget om hvor gode de er til at forske i et bestemt område, sammenlignet med hinanden. Har du nogen forestilling om hvad man kunne gå ind og se på?*

Response: Jamen det skulle nok være noget med hvordan de arbejder med feltet, og hvor mange personer der ligesom beskæftiger sig med det, og hvordan de gør det, fordi en ting er at være en person, der bare sidder inde i et eller andet kælderlokale og bare forsker helt vildt i et eller andet emne, men det kommer aldrig ud af det kontor, så synes jeg det er bedre at man er en større gruppe, som sørger for at komme ud med sin forskning, tager på conference, holder kurser, skriver bøger, laver artikler selvfølgelig, gør opmærksom på sig selv, men jeg tror ikke nødvendigvis at de universiteter, eller det universitet som har mest opmærksomhed også er det som udfører den bedste forskning. Men hvordan man lige skulle måle det, det ved jeg ikke.

Question: *Okay, så har jeg bare lige nogle få små spørgsmål. Først og fremmest så vil jeg bare lige høre om du tilfældigvis ved hvor anvendte de artikler du har med er, f.eks. hvor ofte de er citeret.*

Response: Denne her (2) var jeg inde og kigge, den havde vist fået 15 referencer, det tjekkede jeg faktisk i går.

Question: *Efter du havde fundet den, eller før?*

Response: Efter jeg havde fundet den, denne her fandt jeg i forgårs, og var inde og kigge i går, for at se hvor mange referencer den havde fået og hvad de havde brugt den til. Og den første artikel (1), det har jeg faktisk ikke nogen idé om... den er fra '76, den har også, nej, jeg ved ikke hvad antallet er, men jeg ved at i den bog som er skrevet af forfatterne til artikel (2), de har henvist til f.eks. denne her, men jeg ved ikke præcist hvor mange det er.

Question: *Det sidste jeg har er nogle begreber jeg vil nævne, som du bare skal prøve at svare det allerførste du tænker på, når du hører det. Det behøver ikke være noget langt, det må det gerne være, hvis du har meget at sige om det, men det kan også bare være en kort sætning. <pause> Ny viden?*

Response: Uh oh... arbejde

Question: *Selvfølgelig set i konteksten forskningskvalitet. Du siger arbejde?*

Response: Ja, at man skal følge med i hvad der sker, hvad der kommer af ny viden og forholde sig til det. Så jeg synes det er hårdt arbejde. Men jeg tænker også forbedring og jeg tænker udvikling og perspektivering.

Question: *Så ny viden er i hvert fald noget man kan bruge til noget?*

Response: Ja, spændende, kan man sige

Question: *Tværfaglighed?*

Response: Det er spændende. I mit hovede er det meget anvendelsesorienteret, og det er vigtigt at forbinde forskellige fagligheder.

Question: *Impact - har du hørt om det?*

Response: Ja, men hvis du tænker på impact factor f.eks. så har jeg også hørt om det.

Question: *Hvad tænker du om det?*

Response: Jeg har efterhånden hørt, at det ikke er den bedste måde... jeg tror med det samme du siger impact factor, så tænker jeg "ja ja, men det siger jo ikke hele sandheden", at det er meget afhængigt af hvilken faggruppe du er indenfor, og vores verden, statistik, er meget hårdt ramt med hensyn til impact factor, fordi vores tidsskrifter har ikke særlig høje impact factors, ikke i forhold til indenfor lægevidenskaben. Så impact factor, så

tænker jeg et mål for hvor attraktivt et tidsskrift er at få sin artikel publiceret i, i nogens øjne. Ikke i mine øjne.

Question: *Okay, overhovedet ikke?*

Response: Jeg tænker, nu er jeg lige ved at sidde og skrive en artikel og skulle sidde og vælge et target-tidsskrift med min vejleder, og han snakkede impact factor og jeg snakkede, hvilket tidsskrift jeg synes ville være relevant, hvilke læsere jeg gerne vil have skal komme og læse min artikel. Så jeg tænker impact factor til sidst, jeg tænker det som en bonus, hvis den har en høj impact factor, fordi jeg ved den betyder noget i vores verden, men det er ikke noget der betyder noget for mig.

Question: *Citationer?*

Response: Det tænker jeg nok mere over; at jo flere citationer du har, selvfølgelig alt efter hvordan de citerer dine artikler. Dvs. hvis de citerer dem med, at man gang på gang laver fejl, så er det ikke godt, men det at du har fået citationer til en artikel det gør, at den bliver læst, og det er for mig et bedre tegn på at den er brugbar og at den kommer med en viden man kan bruge til noget.

Question: *Du siger, at det at den får citationer det gør at den bliver læst?*

Response: Nej, at den er blevet læst. Forhåbentlig da.

Question: *Men hvad med det andet, hvis vi nu tager og siger "at den har fået citationer gør at den bliver læst", hvordan forholder du dig til det?*

Response: Jamen det tror jeg også, fordi når jeg sidder og læser en spændende artikel, hvis jeg sidder med en af de her to jeg har med, så kigger jeg næsten altid referencelisten igennem. Dels for at se om der er nogen jeg kender i forvejen, og dels for at se hvilken retning jeg kan gå hvis jeg gerne vil vide mere om et eller andet specifikt i artiklen. Så på den måde er citationer også en ledetråd til hvilke artikler man kan kigge i for at få noget uddybet, så det vil nok også gøre at man... det bider nok sig selv lidt i halen, hvis man først har fået en citation så vil man nemmere få nummer to, de efterfølgende citationer er nemmere end den første, tror jeg.

Question: *Layout af artikler?*

Response: Det er ekstremt vigtigt. Igen indenfor min verden, indenfor statistikverdenen kan vi have artikler som er, selvfølgelig også gamle, men sat op med skrivemaskine, og hvor formlerne er skrevet i hånden. Og vi kan gå helt frem og sige de helt nye, hvor de næsten alle sammen er tospaltede. Dem jeg sidder med her, de to artikler jeg har taget med her, de er ikke tospaltede, de er et-spaltede, hvilket for mig gør at det hurtigt bliver uoverskueligt, end hvis det havde været tospaltet. Artikel (2) er skriftstørrelsen

noget større end i artikel (1), hvilket allerede derved gør at den bliver mere indbydende. Men når man kigger ned over de her artikler kan de godt blive lidt uoverskuelige synes jeg.

Question: *Fordi?*

Response: Fordi lige netop i det her felt, hvor vi har så mange formler med, men problemet med de her tospaltede artikler det er, at tit og ofte så kan formlerne ikke være i én spalte, så en ligning skal deles op på mange linier, hvilket gør det endnu mere uoverskueligt. Men jo flere artikler jeg læser fra det samme tidsskrift, som bibeholder deres layout, jo nemmere er det også for mig at finde rundt i det.

Question: *Så det er faktisk en slags standardisering af layout der er vigtig?*

Response: Ja, for det enkelte tidsskrift

Question: *Og når jeg så siger layout, så kan man jo faktisk tolke det på flere måder. Nu har du valgt een måde, og det er fint, men hvad nu hvis vi i stedet så på layout som opsætning i rækkefølge af tingene?*

Response: Altså i sektionerne?

Question: *Ja, for eksempel*

Response: Jamen der har vi jo typisk indenfor den lægefaglige videnskab, der har man jo typisk den ordening der hedder introduktion, metode og materialer, resultat og diskussion og perspektivering. Og den opbygning gi'r sjældent mening i de artikler indenfor statistikverdenen, fordi du, ja du har introduktionen, og så har du sjældent at det giver mening at lave metode og materiale, fordi det er en teori der skal gennemgås. Det er ikke bare "vi har anvendt denne her statistiske model på denne her", for vi skal først introducere den model som vi skal bruge. Og til denne her slags artikler indenfor statistikken, der giver en anderledes og individuel, til hver artikel, inddeling, det giver mening. Hvorimod til de her registerstudier, epidemiologisk studier, med den her helt klare mening, giver også rigtig god mening, fordi hvis du skal finde noget kan du lynhurtigt gå ind, f.eks. under metode og statistisk analyse og så ved du der skal stå alt det som har noget at gøre med hvilken statistisk model de har anvendt, og så videre. Det er sværere i sådan nogle statistiske artikler og derved så er den røde tråd rigtig, rigtig vigtigt.

Question: *Det sidste jeg har, det har du jo næsten svaret på, det er peer review*

Response: Der tænker jeg, at det er en eller anden black box man sender sin artikel ind i og så ved man ikke hvor mange den kommer frem til og man kan få een reviewer som siger noget helt vildt positivt og man kan få een som nærmest har afvist artiklen før

man overhovedet er kommet videre. At processen er nødvendig, og jeg ved ikke om jeg også vil påstå den er tilstrækkelig, men den er nødvendig for lige netop at sortere nogle artikler fra, hvor man kan sige at metoden den er slet ikke anvendelig. Spørgsmålet er om den er tilstrækkelig i den forstand at alt efter hvor mange reviewere der er på, hvor meget kan de så lige gabe over? Og så er der også alt det med hvor lang tid de har til dette; er det godt eller skidt at de har al den tid de vil nærmest ha' til at læse artiklen igennem, eller er det bedre der er et pres over at man skal være færdig indenfor seks uger, og at du skal have reviewet en artikel for at du selv kan submit en artikel, der er nogle idéer i det, som kunne være fornuftige, synes jeg. Fordi der går så lang tid fra man sender sin artikel ind til at du hører det første gang. For din viden kan faktisk godt være blevet forældet inden.

A.1.2. Interview reference: P02

Date: Sep 06, 2011

Time: 14:30

Question: *Du har taget nogle artikler med, dem vil jeg godt lige be' dig nummerere bare fra 1-2 eller 3 Har du lyst til at fortælle mig noget om hvorfor du har valgt de artikler du har?*

Response: Ja

Question: *Jeg har jo bedt dig tage nogle med, som du vil vurdere som høj kvalitet, hvad er det der er høj kvalitet - der gør det*

Response: Det har jeg jo tænkt over, for hvad er det der gør at noget er høj kvalitet, er det høj forskningsmæssig kvalitet eller er det høj kvalitet for mit arbejde der hvor jeg skal bruge dem, altså er det velegnet til det jeg har skulle bruge dem til for i bund og grund synes jeg det er svært at vurdere en forskningsmæssig kvalitet af et arbejde når man kun ser artiklen. Fordi et er hvad du metodisk skriver du har gjort noget andet er om man så har gjort det man skriver man har gjort. Så jeg vurderer også – selvfølgelig ser jeg også på metodeafsnittet og hvordan de er bygget op. Nu er der jo forskel på kvalitative artikler og kvantitative artikler, fordi i den kvantitative kan man jo sommetider gå ind og vurdere statistikken, men i den kvalitative må man jo vurdere beskrivelsen af processen, for at se på validiteten.

Jeg har valgt to artikler. Jeg har valgt en meget teoretisk metodeartikel som diskuterer, det er den her, af Katrina Voss og ... det er den jeg kalder 2, som er begge to ansat i Norge i Oslo, som arbejder med at kombinere kvalitative og kvantitative metoder i sygeplejeforskningen, om det kan lade sig gøre, og laver studier hvor man kombinerer og så har de en videnskabsteoretisk diskussion af det. Og det synes jeg de gør godt, og det er inspirerende. Det er en inspirerende metodeartikel og derfor synes jeg den er af høj kvalitet. Fordi indholdet er interessant og sætter nogle tanker i gang men også fordi

jeg synes den er rimelig eksemplarisk i sin opbygning. Når man læser den. Så der er både, hvad skal man sige, håndværksmæssigt og indhold jeg nok egentlig kigger på når jeg kigger på kvalitet. Fordi indholdet kan godt være godt, men hvis den er håbløs at læse så kommer man ligesom ikke.. altså

Den anden er et fokusgruppetudie, det er den jeg kalder 1, om seksualitet hos patienter der er behandlet for prostatacancer, den har jo sådan mere faglig tilsnit kan man sige. Det er et fokusgruppetudie der er beskrevet, hvor man har haft 48 patienter i en fokusgruppe, mænd der er opereret for prostatakræft, og har bedt dem om at fortælle hvordan det har været. Primært i forhold til deres seksualitet. Den er også rent metodisk, det jeg kan læse mig til, har de styr på metoden, dvs. at man kan stole, hvis man kan snakke om det inden for kvalitative forskning, på resultaterne. På den måde måde hvor de beskriver både hvordan de har samlet data ind hvad de har valgt, hvad det er for nogle patienter de har valgt, selve fokusgrupperne hvordan de har analyseret deres interviews som siden hen er transskriberet. De finder så frem til nogle interessante, også for mig rent fagligt, temaer, der kan bruges i mit arbejde, og jeg har selv brugt noget af det i en eksamensopgave. I forhold til det her tema. Så der kan man sige, at det meget er resultaterne jeg har brugt og ladet mig inspirere af og bruger i mit arbejde. Så det er. Ja sådan vil jeg sige.

Question: *Det er sådan set egentlig det jeg gerne vil vide om dem. Men hvis vi nu skal snakke lidt mere generelt, så siger du jo selv, at der er forskel på forskningskvalitet og kvalitet af en artikel. Men det er lige netop forskningskvalitet jeg egentlig er interesseret i. Så hvis du skulle prøve at sige noget generelt om det?*

Response: Hvordan jeg vurderer det, eller bare generelt om begrebet forskningskvalitet?

Question: *Hvad tænker du i grunden på når jeg siger forskningskvalitet?*

Response: Jamen jeg tænker, at når man snakker evidensbaseret, både sygepleje og medicin, så skal den behandling man giver skal jo gerne bygge på nogle resultater som man kan stole på, hvis man kan sige det på den måde. Og det er jo primært... tit er det jo artikler man får nogle idéer til nogle nye tiltag fra, sådan er det jo i verden i dag. Det vil sige at man skal kunne stole på at det man konkluderer i en artikel også er det rigtige. Og det er jo forskningskvaliteten det er kvaliteten af den måde man har bygget det studie op på som er refereret i artiklen. Og der synes jeg sommetider det er problematisk at vurdere ud fra de korte beskrivelser der tit er i artikler, af metoden, for at gennemskue om det nu er valide resultater man kan stole på.

Question: *Så når du siger "stole på", så er det faktisk validiteten*

Response: det er validiteten, det er om det er validt. Men det er da klart, at erfaring gør da, at man måske bliver bedre til at gennemskue, altså jo flere artikler man læser. Men jeg mener samtidig, specielt også i den kvantitative del af, at man skal næsten være

uddannet biostatistiker for at kunne gennemskue nogle af de statistiske beskrivelser de har lavet for at kigge på tallene. Som Winston Churchill sagde, han stolede kun på en statistik han selv havde fusket med. Man kan virkelig, i forhold til den slags resultater, jo. Jeg synes jo altså at hvis man kan finde nogen studier hvor man har været inde og sammenligne flere studier og der så når til resultaterne, nu har jeg fuldstændig glemt hvad de hedder... Cochrane

Question: *meta-analyser*

Response: meta-analyser, ja, man skal passe på, at man ikke stoler på det i et enkelt studie, og der er selvfølgelig forskel på kvantitative og kvalitative studier også i forhold til validitet.

Question: *Det kan jo altid være svært, når det handler om kvalitet, at sige noget, hvad skal man sige, at sætte tal på, eller sige noget konkret om det, men hvis man nu alligevel skulle prøve at forestille sig, at man skulle evaluere kvaliteten, f.eks. rangordne eller simpelthen bare sige "det er godt", og "det er ikke så godt", er der noget du tænker der findes? Findes der nogle kriterier?*

Response: Er vi ude i, at der er f.eks. en høj impact factor på et tidsskrift hvor man finder en artikel? det kunne det være, det behøver det ikke nødvendigvis at være det er i hvert fald én parameter man kigger på. Hvis man kan få publiceret en artikel i et anerkendt tidsskrift, så er der en eller anden uskreven garanti for at de reviewere der har været inde over de er dygtige fagfolk, som kan vurdere en videnskabelig artikel og som kan vurdere videnskabeligheden af det arbejde.

Question: *Mener du at der er en sammenhæng imellem høj impact og så den peer review process der ligger bag?*

Response: Ikke nødvendigvis, man må formode, men der er så meget... jeg ved ikke hvor det kommer hen alt det jeg sidder og siger, men der er jo meget sammenspisthed...

Question: *du behøver ikke være nervøs for at det kommer nogen steder hen*

Response: nej, men det jeg sommetider hører nogen sige det er, vi sender til det tidsskrift, for der kender jeg editor og ved hvem der sidder som peer reviewer, dem har jeg arbejdet sammen med på et tidspunkt, underforstået, så ser vi måske lidt mere mildt på hinandens studier. Så der er ikke nødvendigvis nogen sammenhæng, men der burde være, hvis vi sådan i den rigtig gode verden, hvor vi alle er gode og ordentlige, så burde der være en sammenhæng. Selvfølgelig, det der med at henvise til er jo ikke nødvendigvis... det kan jo også være fordi det er et interessant tema som lige er oppe i tiden, som alle gider at henvise til, men principielt burde det jo være sådan at de mest estimerede tidsskrifter ligesom er dem man bygger studier videre på, og når man henviser til noget er det jo fordi man henviser til andre har lavet noget der ligner, og jeg

laver lige en ekstra lille del, eller...

Men altså, det er en eller anden form for en pseudo-verden indimellem også, det her, tænker jeg. Og jeg tænker lidt at det er det samme indenfor kvalitativ forskning, fordi, der har man jo ikke på samme måde, altså man kan sige indenfor kvantitativ forskning kan man jo, hvis man går tilbage og finder alle tallene og finder ud af hvordan studiet er bygget op kan man jo på en eller anden måde genbearbejde resultaterne igen. Kvalitativ forskning har jo lidt mere, er jo lidt mere løs i koderne, selvom man ikke kan sige det på den måde, men qua at det jo er relationer og processer, det kan vi jo ikke genlaves på den måde, så der er jo en anden form for validitet og generaliserbarhed, men også en anden form for at gennemskue om det lyder plausibelt om det her studie er lavet korrekt.

Question: *Du har sagt to ting som jeg gerne vil have dig til at uddybe lidt, den ene ting du nævner er studier, der er meget oppe i tiden, altså en form for aktualitet.*

Response: ja

Question: *Mener du det er en decideret kvalitet ved et studie at det behandler noget, der er aktuelt?*

Response: Nej, det mener jeg ikke. Ikke nødvendigvis da, for det er jo ligger jo tit i slipstrømmen af noget der er politisk korrekt eller der ligger penge i, i forhold til fonde eller, ja, kan der jo godt ligge en masse knap så solide studier, fordi det måske er nemt at få penge lige nu, eller fordi det er smart for en selv som forsker at få sig promoveret på et felt som har politisk bevågenhed. Nej, jeg mener ikke der er en sammenhæng. Jeg vil ikke engang sige tværtimod, men sommetider, jo, det var i forhold til noget med peer review, altså noget der er oppe i tiden, er jo tit noget der bliver gentaget i pressen, så nej, der er ingen nødvendigvis, tværtimod kan man måske sige...

Question: *Så ligger der jo lidt implicit i det du sagde med impact factor: henvisninger, at der ligger en eller anden forbindelse imellem det at der bliver henvist til noget og det at man opfatter det som noget der har en eller anden form for kvalitet. Oplever du det også sådan eller er det noget andet?*

Response: Nej, ikke nødvendigvis, men der er alligevel inden for, tror jeg, inden for hvert speciale eller hvert forskningsfelt nogle gode og nogle onde. Altså, noget der er bedre at referere til end noget andet, altså nogle uskrevne, jeg vil ikke sige vedtægter, men altså, hvem der er gode og hvem der ikke er gode. Jeg tror at man, altså dels ligger der jo i at hvis man skriver til et tidsskrift, så vil de jo gerne have at man refererer til tidsskriftet selv, det får man jo tit at vide.

Question: *Er det noget du har oplevet selv?*

Response: ja, at de undrer sig over, at man ikke har refereret til tidligere arbejder i eget

tidsskrift?

Question: *Ja*

Response: Så på en eller anden måde kan man sige at det der med impact factor det er jo også igen et eller andet pseudo. Det er nemmere at få optaget en artikel i et tidsskrift hvis man har et par referencer der refererer til tidsskriftet selv. Nu kom jeg lige fra hvad det var vi snakkede om, jeg fik ikke forklaret mig ordentligt

Question: *Du snakkede noget om en slags tradition for hvad man citerer.*

Response: Ja, det tror jeg også der er, der er de der grand old men indenfor, ja det er mest smerteforskning jeg har været indenfor, der er helt klart nogens arbejder der er fornuftige at citere og bygge videre på.

Question: *Fordi det gør det nemmere for en at komme igennem?*

Response: Det gør det nemmere for en selv at få noget publiceret, og nogen vil også sige at nå ja, hvis de refererer til dem og bygger videre på deres arbejde så er det nok fordi det også er solidt det her.

Question: *Lidt tricky at svare på måske, men du har selv nævnt at der ikke nødvendigvis var en sammenhæng imellem selve artiklen og så det arbejde der lå bag, og der har du jo så foregrebet begivenhedernes gang lidt, men er der en eller anden relation, vil du mene, eller er det bare fuldstændig uden?*

Response: Nej nej, der er helt klart en relation, altså imellem artiklen og det der ligger til grund for artiklen, det har jeg ikke fantasi til at forestille mig at der ikke er. Men til gengæld tror jeg som verden er, med at det i dag tæller helt enormt med en publiceringsliste så kan jeg være bekymret for om man i nogle tilfælde begynder piller lidt ved nogle resultater eller udelader nogle ting, fordi... jeg havde en diskussion med Conni, fordi jeg har lavet et studie hvor vi simpelthen ikke fik et eneste positivt resultat ud af, og så siger jeg til Conni, hvor har man tradition for at publicere negative resultater, og det har hun jo ledt lidt efter, og vi har ledt, og vi har ikke rigtig kunne... og hvor er de henne alle de der, hvor de ikke viser noget? Eller drejer man det så lidt i en retning, eller hiver man kun de der få lave p-værdier frem, der har en tendens til at være statistisk signifikante, eller hvad han skriver, ham statistikerer jeg arbejder sammen med, ja, jeg er sådan lidt almindelig skeptisk, når jeg ser nogle ting, om det nu er så godt, eller... men altså, sund fornuft og så tanker folk har gjort sig er jo altid interessante, men jeg ville aldrig ud fra en artikel, hvis jeg nu var behandler, så sætte en ny behandling i gang ud fra én, der vil jeg se mere. Sådan har jeg det.

Question: *Ja, så har jeg en lille smule mere; nogle begreber jeg vil nævne, og hvis du så bare siger det første du tænker på. Selvfølgelig gerne i den kontekst.*

Response: Ja

Question: *Ny viden?*

Response: Hold da kæft... ny viden, jamen det er vigtigt. At producere ny viden.

Question: *Tværfaglighed?*

Response: Det er virkelig vejen frem. Det er det vi skal satse på.

Question: *Impact*

Response: Det er den jeg er sådan en lille smule loren ved, fordi den burde sige noget om kvalitet, men det er jeg ikke sikker på den gør.

Question: *Citationer?*

Response: Ja.. for hvis skyld? Det er igen.. er det fordi det er politisk korrekt, eller er det fordi det er taktisk, eller er det solidt arbejde?

Question: *Layout af artikler?*

Response: Der har også en stor betydning, selvom det ikke burde have det. Vi bombarderes af så mange ting i dag, så at det visuelle også har en betydning. Er det dét du tænker på når du siger layout?

Question: *Ja, netop*

Response: Ja, det har også en betydning. Når du siger layout så tænker jeg faktisk mere på det jeg siger i forhold til at lave en eksemplarisk artikel, altså at lave en læsevenlig, altså struktureret artikel, det betyder noget. Nu sidder jeg også lige som reviewer og er lige i gang med en artikel som altså bare er så elendig, og altså, og det er manglende sammenhæng og hvor jeg tænker at jeg orker næsten ikke, og det er sådan en artikel hvor jeg tænker, at hvis den var gået igennem og jeg ville sidde med den i et tidsskrift, så ville jeg lægge den fra mig. Selvom det måske var nok så interessant.

Question: *Og det sidste jeg har er også peer review.*

Response: Ja, i ordets betydning er det jo en fantastisk god ting, at man finder nogen som er ligestillede og har forstand på det felt til at gå ind og vurdere. Så på den måde er det jo en god ting, for det burde være de ypperste og dem der ved mest om emnet, der er dem der går ind og vurderer hinandens arbejder. Men igen synes jeg at der ligger meget andet i det også, at det er vennetjenester og kammerateri.

Question: *Ja, det var så det. Hvis du har andet du gerne vil tilføje, hvis der er noget der*

er usagt, så må du selvfølgelig gerne det.

Response: næ, ikke andet end at vi jo ikke kan leve uden den måde som det er nu, altså den måde at publicere på, men jeg synes bare man skal have sund fornuft og en lille smule, ikke negativitet, det er på ingen måde det jeg udtrykker, men man skal ikke tage råt for usødet, men skal lige have sund fornuft med.

A.1.3. Interview reference: P03

Date: Sep 20, 2011

Time: 16:00

A.1.4. Interview reference: P04

Date: Sep 09, 2011

Time: 12:30

Question: *Vil du fortælle mig noget om hvorfor du har valgt at tage de her to artikler med?*

Response: Den første her, der tænkte jeg over, hvad jeg egentlig mener er høj kvalitet, og det er forskellige kriterier synes jeg, det kan godt være et lille studie i et lille tidsskrift, men som er lavet på en god idé. Nu endte jeg alligevel med to forholdsvis store studier, faktisk i det samme tidsskrift, et meget velanset tidsskrift, det var sådan det blev, men det er ikke ensbetydende med at det er en kvalitet i sig selv. Det første studie synes jeg har høj kvalitet fordi det er et studie som besvarer et klinisk spørgsmål som er meget vigtigt, som betyder noget, det her studie fik umiddelbart betydning for patientbehandling, det gør at vi ændrer behandlingen og det kommer en masse patienter til gode. Så på den måde synes jeg det er en høj kvalitet, og det er udført tilpas omhyggeligt og stort nok, altså sådan at de faktisk kan besvare alle spørgsmålene de stiller. Så det er i mine øjne god kvalitet. Der er selvfølgelig også det ved det at det er et hamrende dyrt studie, det er ikke sådan noget man lige laver som ph.d. studerende, men ikke desto mindre så er det høj kvalitet, synes jeg.

Det andet er et lidt mindre studie som nok mest af alt er landet i det her tidsskrift fordi det blev et interessant resultat, men også fordi det er godt lavet. Her har man også stillet et væsentligt spørgsmål, det her går på noget med genotyper og risikoen for at udvikle åreforkalkning og det er sådan noget der kan være lidt svært at undersøge, og de har så valgt ikke at undersøge det på de typiske outcomes som er blodpropper, de har valgt det på en mellemliggende outcome som er åreforkalkning målt ved ultralyd i halse. Det har de sådan set gjort meget systematisk og godt, og det de så ovenikøbet har gjort er at kigge på noget interaktion imellem de gener her og nogle bestemte fedtsyrer. Og ikke nok med det, det kan man sige der er egentlig mange der kigger på interaktion, men det her var godt lavet fordi det hænger godt sammen i teorien, det var meget velunderbygget

hvorfor de valgte at gøre det på den her måde, og det er det jeg mener der er god kvalitet her. De har tænkt meget over det design før de gik ud og lavede studiet. Det er det der har endt med at gøre det godt. Og så er det ovenikøbet også meget pænt præsenteret. Det betyder også noget, synes jeg. Det kommer ud i et lækkert design med nogle gode figurer, som er til at gå til.

Question: *Hvis vi nu snakker forskningskvalitet helt generelt, og ikke med udgangspunkt i de her to artikler, hvad tænker du så? Er der noget du vil sige er afgørende for at noget er høj kvalitet?*

Response: Jeg synes det er svært at sige een ting, det er flere ting der skal være opfyldt for at det er god kvalitet. Man kan sige helt overordnet så er det en god kvalitet at man kan besvare det spørgsmål som man har tænkt sig at undersøge i et studie. Hvis man ikke kan besvare spørgsmålet fordi man ikke har lavet en styrkeberegning eller et eller andet, jamen så er det ikke god kvalitet, så har man ikke tænkt sig om på forhånd.

Question: *Og det ser man?*

Response: Det ser man, så sker der typisk det, at folk begynder at kigge på noget lidt andet end det de oprindeligt ville, når de finder ud af at de ikke kunne besvare spørgsmålet, så bliver det lidt derefter. Og man kan sige noget andet som er lidt i forlængelse af det, så nogle gange så bruger vi alle sammen for lidt tid på at forberede studier, altså tænker for lidt over hvor hvilke variable vil gerne vil have med, hvilke problemer der er med selektionsbias og forskellige ting. Det er nogle af de ting, hvor det er god kvalitet hvis det er velovervejet fra starten af, at man har valgt det ud inden man begynder at indsamle og kigge på sine data. Det er nogle af de grundlæggende ting som bør være i orden. Så kan man sige at god kvalitet kan være forskellige typer af design. Indenfor vores verden synes vi typisk at den bedste kvalitet er randomiserede studier, så det er klart at hvis man vil teste et eller andet, så ser man jo helst et randomiseret studie i en pæn størrelse, der er veldokumenteret, det er god kvalitet.

Question: *Så evidensniveau er beslægtet med kvalitet?*

Response: Ja

Question: *Vil du sige der er en direkte sammenhæng?*

Response: Nej, det vil jeg egentlig ikke. Jeg mener godt du kan lave et epidemiologisk studie, retrospektivt, som i sigens natur ikke vil have et meget højt evidensniveau, men som alligevel godt kan være udført i en meget høj kvalitet. At man har lavet det så godt som man overhovedet kan i den type design.

Question: *Men der er i hvert fald en eller anden sammenhæng imellem dem?*

Response: Ja, det er klart, at hvis vi skal ud og tage beslutninger på det forskning der er lavet, der går vi efter det der har højest evidensniveau og det er typisk de store randomiserede studier. På den måde kan man sige der er en sammenhæng.

Question: *Hvis man skulle ud og foretage en evaluering af forskning og forskningskvalitet, det kunne være to universiteter man ville sammenligne, eller det kunne være forskningsgrupper, forskere, det kunne også være enkelte artikler. Er der noget man kunne se på, for at afgøre hvem der leverer den højeste kvalitet? Nu har du sagt evidens, er der andet man kunne se på?*

Response: Noget af det man formentlig ser på, jeg ved ikke rigtig, man kan se på hvilke tidsskrifter man publicerer i, hvad er impact factor i de forskellige tidsskrifter, der kan man jo formentlig finde en gennemsnitlig impact factor pr. universitet, det kan man jo godt bruge i en eller anden udstrækning, som kvalitetsstempel. Jeg synes det er svært at gøre op, jeg ved ikke hvordan man skulle gøre det op, men en af de ting som vi bør gå mere op i er om det betyder noget, det forskning, om det ændrer noget i den måde andre agerer på, om det har et potentiale til at have en indvirkning på sundhedsøkonomien eller den generelle sundhedstilstand.

Question: *Har du nogen forestilling om hvordan man på et enkelt studie kunne gå ind og se på hvilken effekt det har haft?*

Response: Du tænker efterfølgende?

Question: *Ja*

Response: Jamen det kommer lidt an på, man kan jo sige at i den her type studier som er så kanon store her, der kan du gå ind og se at efter det studie her, har guidelines ændret sig. Det står direkte at det skyldes i stor udstrækning som dette her, så der vil du kunne se at det har betydet utrolig meget for både sundhedstilstand, sundhedsøkonomi og den måde man behandler, men det er jo de færreste studier der er sådan, og de andre vil det være mere indirekte, hvor man kan sige at det her forskning har været med til at nogle andre har kunnet bygge ovenpå og i sidste ende er det noget der flytter et eller andet.

Question: *Men vil du mene, at hvis og så fremt, et studie bliver inkluderet i en guideline, så er det fordi det studie har en eller anden form for klinisk relevans og dermed en eller anden form for kvalitet?*

Response: Ja, det vil jeg sige, så er det fordi det betyder noget.

Question: *Nu nævner du selv impact factor og tidsskrifter, og de to artikler du har taget med er jo faktisk fra NEJM, som jo nok er det tidsskrift der har den højeste impact factor, i hvert fald indenfor det medicinske område. Ser du en sammenhæng imellem det? Altså*

oplever du en sammenhæng imellem det, eller er det mere noget du hører? Du giver udtryk for en tøven, synes jeg.

Response: Der er ingen tvivl om at der er utrolig mange af de studier der bliver publiceret deri der er kanon gode. Det mener jeg faktisk de er. Man kan sige det er to typer af studier der bliver produceret heri, det er de rigtig store, der koster mange millioner at udføre, som er dem der besvarer nogle klinisk vigtige spørgsmål, de bliver publiceret heri. Det er sådan set også fint nok. Og de andre som bliver publiceret heri, det er nogen som viser nogle ekstraordinære resultater, og der er jo nogen som kalder de her tidsskrifter "journal for ikke reproducerbare resultater", og det er der jo nok nogen sandhed i, fordi nogle gange er det jo tilfældighedernes spil at man lige har det her studie, som har vist noget helt ekseptionelt, og man kan ikke reproducere det fordi det har været lidt tilfældighedernes spil, de har været så heldige at få sådan nogle resultater. Så det er klart på den måde kan man blive misvejledt af sådan nogle tidsskrifter nogle gange.

Question: *Men du oplever at der er en eller anden form for sammenhæng?*

Response: Jeg tror da de fleste vil sige at generelt set har de en høj kvalitet. Også fordi det kræver noget at få sine ting publiceret deri. Jeg tror de har en lidt større revision man skal igennem, end så mange andre tidsskrifter. De kigger lidt mere ind i metode og statistikker, om det er i orden lavet. Jeg har da også set eksempler på noget der går igennem [som ikke er godt, red.], men jeg tror da det hører til sjældenheder.

Question: *Så har jeg kun en lille del mere, jeg har et enkelt spørgsmål og en brainstorming. Nu har vi snakket meget om selve publikationerne, men der ligger også noget forskning bag ved, nogle studier. Mener du der er en sammenhæng imellem dem, eller kan man ligeså godt producere en god artikel på baggrund af et dårligt studie?*

Response: Jeg tror det hænger meget godt sammen. Vi ved jo allesammen godt at det er utrolig meget lettere at skrive en god artikel hvis du har nogle klare resultater, og de klare resultater får du bedst hvis du har designet dit studie godt på forhånd. Så bliver det nemmere at skrive og så giver det et resultat der er meget nemmere at kommunikere ud til omverdenen. Hvis du har et dårligt studie og skal til at skrive om det jamen så bliver det tit noget væven omkring en masse ting man skal tage højde for og lige tage med i overvejelserne. Det bliver svært at kommunikere ud. Jeg synes det giver en lidt artikel også, så det tror jeg hænger nogenlunde godt sammen.

Question: *Så har jeg som sagt en brainstorming-session. Det går ud på at jeg nævner et begreb, og så vil jeg gerne have du siger det første du tænker på. Om det så er kort eller langt, positivt eller negativt, det er helt op til dig. Det første ord er ny viden.*

Response: Innovation

Question: *Tværfaglighed?*

Response: Problemer. Skal jeg uddybe det?

Question: *Ja, det må du meget gerne, hvis du vil*

Response: Det første der falder mig ind er problemer, og det er fordi i et studie jeg kører, hvor jeg skal samarbejde med flere forskellige specialer, det bliver lidt problemfyldt, fordi der er mange hensyn der skal tages. Det var lige det, der faldt mig ind. Der er også mange positive ting ved det.

Question: *Hvis vi nu skal se det i sammenhængen forskningskvalitet, hvad tænker du så?*

Response: Jeg tror for det meste det er godt, jeg tror bestemt også det kan være skidt. Fordelen ved at det ikke er tværfagligt er, at man snakker samme sprog, problemet er nogle gange, at hvis det bliver for tværfagligt så vil man gerne hive det i hver sin retning, og så bliver det noget underligt noget midt imellem som man ikke rigtig kan bruge til noget.

Question: *Impact har du snakket lidt om, men hvad tænker du?*

Response: Jamen jeg tænker store tidsskrifter. Jeg kan ikke engang præcis huske definitionen, det er vist noget med hvor mange der er citeret i ens artikel i en måned, tror jeg.

Question: *Det er heller ikke så vigtigt. Nu har du næsten nævnt det næste; citationer?*

Response: Det første der lige falder mig ind det er sådan set referenceprogrammer. Så det er jo lidt noget andet.

Question: *Du har ikke rigtig nogen holdning til hvorvidt de spiller en rolle i forhold til en vurdering af hvorvidt en artikel er god eller ej?*

Response: Tænker du på om jeg mener den er god eller dårlig fordi den er blevet citeret mange gange?

Question: *Enten det, eller om man ud fra mængden af citationer kan se om den måske er god eller ej. Om der er en eller anden sammenhæng?*

Response: Jeg tror der er en sammenhæng imellem at hvis den er citeret mange gange, så er det fordi den er interessant. Og det kan så både være i positiv og i negativ sammenhæng, fordi det er klart at en artikel også kan blive citeret mange gange for noget negativt. Men for det meste vil jeg tro at de bliver citeret mange gange fordi det er noget der betyder noget.

Question: *Så har jeg et begreb der hedder layout af artikler*

Response: Layout, når du siger det så tænker jeg der er mange forskellige tidsskrifter, der har hver deres måde de gerne vil have et layout på. Og det kan jeg godt forstå, og nogle gange synes jeg det er lidt ærgerligt, f.eks. vil de ofte gerne have det er i sort/hvid. Jeg kan godt forstå hvorfor man ikke vil have alle mulige spraglede, men nogle gange kan man altså gøre det lidt mere spiseligt hvis det kommer i farver. Jeg synes nogle gange layoutet er lidt for konservativt, det kunne godt være lidt mere lækkert.

Question: *Har det noget at sige i forhold til hvordan du opfatter en artikel?*

Response: Jamen det har det. Jeg kan klart bedst lide at læse en artikel hvis der er nogen rigtig gode figurer, nogle gode illustrationer, og der må gerne være farver på, der gør det mere spiseligt. Det synes jeg helt klart.

Question: *Det sidste begreb jeg har er peer review*

Response: Ja ja, jamen der tænker jeg selvfølgelig at det er noget der er læst igennem af ligesindede tror jeg nærmest det betyder. At der er andre der har reviewet ens artikel inden den bliver optaget. Og det synes jeg er en god ting, at man har det. Det typiske jeg forestiller mmig er, at når man sender noget ind finder man nogle andre forskere, der har forsket i det samme, og får dem til at læse det igennem. Og det synes jeg er en god kvalitetskontrol, at der er nogen der ved noget om det, der læser det, før det bliver publiceret til resten af verden. Det er også en måde at få feedback på, til hvad man kan ændre og gøre bedre.

Question: *Så har jeg bare to små spørgsmål til noget du har nævnt, eller, det ene er til noget du har nævnt. Du nævnte at man kan citere noget for noget godt og noget skidt, og du nævner også at man måske ofte citerer noget for noget dårligt...*

Response: Nej, jeg synes ikke man ofte gør det, men har man en eller anden artikel, som er meget kontroversiel, og som er blevet beskyldt for at være et dårligt design, f.eks., så vil man tit se at den bliver citeret mange gange, som et studie der afviger fra de andre. Jo, det har jeg set i nogle forskellige sammenhænge. Og jo det er selvfølgelig ikke et kvalitetsstempel for den artikel.

Question: *Okay. Det andet jeg tænkte på var; har du helt tilfældigvis været inde og se hvor tit de to artikler du har med er citeret?*

Response: Nej, det har jeg ikke

Question: *Har du en fornemmelse af det?*

Response: Med det mener du hvad impact factor er på de her to?

Question: *Nej, siden de er blevet udgivet, hvor mange gange er de blevet citeret? Det er selvfølgelig svært at svare på, hvis man ikke har været inde og kigge.*

Response: Den her er fra 2004, nej, jeg tør ikke engang at skyde på det.

Question: *Det er også bare helt i orden. Det var sådan set det så. Hvis du har et eller andet du synes du mangler at sige, så må du selvfølgelig meget gerne det.*

Response: Nej, det er der ikke.

A.1.5. Interview reference: P05

Date: Sep 16, 2011

Time: 10:00

Question: *Jeg har bedt dig tage de her to artikler med, som du har vurderet til at være af særlig høj kvalitet. Vil du fortælle mig noget om hvad du vurderer som høj kvalitet ved dem?*

Response: Ja, jeg tager artikel 1, når jeg tager fat i den så er det fordi det er en af de artikler som min ph.d. bygger videre på. Det er et stort follow-up studie fra USA og hvorfor synes jeg det er høj kvalitet? Det er svært at sige... det virker vældig gennemført, at de har styr på de forskellige eksponeringer og variable og er ret stringente i deres analyser.

Og det er en artikel som har åbnet op for spørgsmål om effekt af fisk og fiskeindtagelse på atrieflimren som er det jeg beskæftiger mig med. Så det var den første store epidemiologiske undersøgelse, der rapporterede noget om det.

Question: *Så det har åbnet op for nogle spørgsmål omkring emnet?*

Response: Ja. Og man kan sige de har lavet sig nogle overvejelser i forhold til eksponeringen, og de deler også fisk op i om det var fisk eller sådan noget stegt fisk, og der så de forskellen imellem de to grupper og argumenterer meget godt for hvorfor de fandt det.

Question: *Hvis jeg må prøve at summere lidt op, så siger du noget om stringens i deres metoder og at de åbner op for nye forskningsemner?*

Response: Ja

Question: *Hvad med den anden?*

Response: Jamen der sker jo flere ting i den artikel her. Jeg synes den er meget velformuleret, den er præcis i beskrivelse af hvad den undersøger og den har dels den styrke at det også er et stort follow-up studie, et studie hvor man anvender fiske-, altså nogle vævsprøver, nogle fedtvævsprøver, og data omkring det, hvilket er sjældent at man har

taget sådan noget med i studiet, så det er sådan set en enestående studiemulighed man har udnyttet der, og så har man lavet et designmæssigt elegant design, som er sådan et case-kohorte design, hvor man udnytter at man ikke behøver lave fedtvævsprøver på alle deltagere. Man har lavet en kontrolgruppe, der repræsenterer kohorten, og så synes jeg den er meget velskrevet og har mange ting med, men er ret præcis i hvordan den får forklaret hvad der er sket, og hvorfor man har gjort de forskellige ting, hvorfor man har udvalgt de forskellige ting, hvorfor man har lavet de statistiske analyser man har valgt og hvorfor man har taget nogle ting med og ikke justeret for andre ting. Så på den måde at man også lidt lærer af den type kommentarer og tænker "okay, det er derfor de har gjort det."

Question: *Nu skal jeg lige være sikker på jeg har forstået det rigtigt, men jeg hører lidt at du siger, at den fremgangsmåde de har haft har været, nu lyder det måske lidt negativt, det mener jeg ikke; kreativ? Altså at det ikke er en fuldstændig standard, eller misforstår jeg?*

Response: Det misforstår du, det er egentlig det man kalder et case-kohorte design, og det er egentlig re nyt design, men rigtig intelligent design. Så det er ikke på den måde smart, som i at det ikke er troværdigt. Det er smart som i at...

Question: *Det var heller ikke det jeg mente, jeg mente mere at de måske havde gjort noget der lå lidt udenfor det normale, men det er det så heller ikke?*

Response: Nej, det de har gjort, som er lidt ud over det normale er, at de har haft nogle fedtvævsprøver som man har kunnet analysere, og så har man valgt at gøre det med det her velegnede design.

Question: *Okay, jeg skulle bare lige være helt sikker på jeg havde forstået det. Nu har du jo faktisk flere gange nævnt f.eks. follow-up studier, case-kohorte, det andet sagde du var et epidemiologisk studie, det får man til at tænke på evidens. Er det noget du sætter i forbindelse med forskningskvalitet?*

Response: Ja, det er det helt klart. Det kommer selvfølgelig an på spørgsmålet man stiller sig i forskningen. Men det jeg er mest indenfor det er, er der evidens for det ene og det andet? Og i den sammenhæng er det selvfølgelig rigtig vigtigt. Men jeg vil sige at der selvfølgelig også er andre forskningsområder hvor det ikke er en tilgang man kan [bruge].

Question: *Okay. Så har vi snakket om de her to artikler lidt, om hvad det er der gør, at de har en eller anden kvalitet. Hvis vi nu skal se mere generelt på forskningskvalitet, hvad tænker du på når jeg siger det?*

Response: Forskningskvalitet er jo noget der bringer relevant og brugbar viden til et højere niveau, som kan man sige, gør nytte på den måde at det giver mening overhovedet at høre om resultaterne. Det vil jeg sige, og så selvfølgelig at det er gennemskueligt hvad

der er sket, og at egentlig også at man kan forstå hvad der er sket.

Question: *Så det er egentlig noget med opsætningen af artiklen, at man har skrevet hvad man har gjort og hvorfor man har gjort det?*

Response: Ja, og den måde man har valgt at gøre det, designet, ser jeg som en stor del af forskningskvaliteten.

Question: *Du har ikke mere til det?*

Response: Nej

Question: *Helt i orden. Hvis man skulle forsøge objektivt at evaluere kvaliteten af forskning; kan man det? Og så fremt du mener man kan, har du nogen idéer om hvordan?*

Response: Hvordan skal man bedst vurdere det? Så vil jeg sige, jeg vil gerne høre hvad et bredt panel af sammensatte forskere, 100 forskere, der scorede dem i forhold til relevans og brugbarhed og hvordan det bibringer ny viden. At vurdere på baggrund af hvor mange andre artikler der citerer den artikel, det er selvfølgelig et mål, men der bliver bare gerne nogle artikler der bliver ikoner, som man skal have med, for det skal man bare, selvfølgelig vil der være nogen af dem der har høj værdi, men det er lidt ligesom... det er selvfølgelig meget afhængigt af emnet, det er jo kun væsentligt indenfor det man selv tænker.

Question: *Nu nævner du citationer, det går jeg ud fra fordi dem kender du som en slags kvalitetsmål...*

Response: ...ja, en måde at gøre det op på.

Question: *Har du tilfældigvis selv været inde og se hvor citeret de to artikler du har med er?*

Response: Nej

Question: *Og du har heller ikke nogen fornemmelse af det?*

Response: Nr. (2) er helt ny, så det er den ikke, men den første, jo, den er selvfølgelig citeret i alle efterfølgende artikler om fiskeindtag.

Question: *Så den har du faktisk ofte set på referencelister?*

Response: Ja, den er oftest med, ja, den ser jeg ofte.

Question: *Okay. Så har jeg egentlig kun en kort brainstorming, hvor jeg nævner nogle begreber, og så vil jeg gerne have du kort svarer hvad du tænker når jeg nævner dem, selvfølgelig*

i den kontekst, der hedder forskningskvalitet. Det første er ny viden.

Response: Det tænker jeg er meget relevant, det har stor betydning for om det er noget der er værd at høre på.

Question: *Tværfaglighed?*

Response: Ligemeget

Question: *Impact?*

Response: Det er også, kan man sige, jeg tænker sjældent forekommende men væsentlig del, på den måde at det alligevel er relativt sjældent man har noget forskning der har stor impact, hvis man har, så er det selvfølgelig af høj værdi og derved også kvalitet. Så med ordet impact forstår du at en forskningsartikel gør en forskel.

Question: *Citationer har du fortalt om. Men hvad med layout af artikler?*

Response: Jamen jeg kan godt lide grafer. Dem forstår jeg rigtig godt.

Question: *Vil du mene det giver en øget kvalitet hvis en artikel har det?*

Response: Ja, jeg vil hellere læse artikler med grafer.

Question: *Peer review?*

Response: Jamen helt væsentligt. På den måde at en ikke-peer reviewed ting anser jeg bare som et interessant journalistisk aspekt, men ikke noget jeg er interesseret i at læse. Med mindre det er, jeg ved egentlig ikke hvordan ledere eller editorials er i tidsskrifterne, men det skal være et peer-reviewed tidsskrift for at jeg gider at læse det i detaljer, bl.a. fordi der bare er så meget information, og det der ikke kommer de steder er typisk vinklet på en måde som gør at det er irrelevant.

Question: *Når du siger, at der er så mange informationer, så går jeg ud fra det er underforstået, at man ikke kan nå at læse det hele?*

Response: Ja, netop

Question: *Og forstår jeg dig så ret, at peer review er med til at sortere noget fra, som er mindre væsentligt?*

Response: Ja

Question: *Så var det egentlig det jeg havde, men der var en anden der nævnte et begreb som jeg egentlig også gerne vil have et meget kort svar på, eller hvordan du forholder dig til;*

klinisk relevans?

Response: Der vil jeg sige 50-50, det er sådan at nogle forskningsting har klinisk relevans på den måde at det er noget som kan være relevant at vide når man arbejder som læge, f.eks., men det vedrører den forskning som er i slutningen af fødekæden, kan man sige, det er meget tæt på patient-behandlingen, "hvad bliver de kommende guidelines", det er helt derude hvor man er nået så langt så spørgsmålet egentlig handler om relevante behandlingstiltag, hvorimod langt det meste forskning ligger langt forud for det, hvor det ikke har nogen klinisk relevans for dem der aktuelt arbejder med patientbehandling. Så jeg ser klinisk relevans som når dem som arbejder med kliniske problemstillinger i hverdagen mener at de har interesse i at læse det. Der er meget grundforskning som langt ude i fremtiden kan have klinisk relevans. Jeg ser det ikke som klinisk relevans før det.

Question: *Det er meget interessant synes jeg. Jamen det var det så, medmindre der er noget du føler du mangler at sige?*

Response: Nej

A.1.6. Interview reference: P06

Date: Sep 09, 2011

Time: 11:30

Question: *Vi har nummereret dine artikler, og jeg har bedt dig tage dem med baseret på at du vurderede dem til at være høj kvalitet. Har du lyst til at fortælle mig noget mere specifikt om hvorfor du har valgt netop de her to artikler?*

Response: Jamen det kan jeg godt. Vi kan jo tage dem fra en ende af, nummer (1) her til at starte med, det er jo sådan at man kan vurdere forskningskvalitet på mange måder og jeg synes det der er godt ved den her artikel (1), det er... den omhandler et område omkring når man skal fjerne elektroder på patienter som har indopereret en avanceret pacemaker. Og der er lavet rigtig mange studier som viser at det kan man gøre uden de helt store problemer. Problemet er bare, at alle de studier er fra nogle enkelte få centre hvor de ikke har særlig stor styrke til at udtale sig om det de egentlig udtaler sig om. Det som de forfattere så har gjort i det her arbejde er så at de i stedet går ind og kigger på en database fra industrien og har så fået et multinationalt, både fra USA men også fra Europa, lidt input derfra, og der viser det sig så at der rent faktisk på verdensplan er rigtig mange, eller i hvert fald en del, patienter der kommer galt afsted med at få fjernet de her elektrode. Så den påpeger nogle ting som er meget, meget vanskelige at påpege ved de opgørelser der ellers i øvrigt er lavet fordi man har gjort det på en helt anden måde med en anden tilgang via industrien og det synes jeg egentlig er ret godt tænkt og den er også med til at stille spørgsmålstejn ved om den måde man anskuer

det på idag er i orden eller om man skal undersøge det nærmere, så det er egentlig en hypotesegenererende artikel mere end det er en der oser af statistik eller noget som helst andet, det er en artikel som rent deskriptivt beskriver en problemstilling som i hvert fald kan sætte nogle andre ting igang og det synes jeg må på en eller anden måde definere god forskningskvalitet; altså noget som både påviser en relevant problemstilling og stiller nogle spørgsmål som man så kan arbejde videre med. Og i noget af det jeg selv skal arbejde videre med er det en artikel som jeg gentagne gange er faldet tilbage til fordi den stiller nogle relevante spørgsmål som jeg synes jeg kan bruge til noget. Også set ud fra et klinisk synspunkt, set ud fra lægens synspunkt. Så jeg synes egentlig det er en god artikel, den er godt tænkt og de registre man har brugt er nogle lidt andre registre end ellers ser i andre opgørelser, så det synes jeg er en god artikel. Skal jeg fortælle lidt om den anden artikel?

Question: *Ja, gør det*

Response: Det er jo sådan lidt en klassiker, det er også noget jeg har siddet og arbejdet lidt med nu her, det er en der omhandler ballonbehandling ved patienter med blodpropper i hjertet. Et dansk studie, DANAMI-2 studiet, som egentlig primært går på at man undersøger om det kan betale sig at flytte patienter fra henviste sygehuse til et hjertecenter for at lave ballonbehandling eller om man i stedet skal give noget blodfortyndende medicin bare på stedet, hvis de har en stor blodprop i hjertet. Og det der er elegant ved studiet her, det er for det første, at det er en meget klinisk relevant problemstilling, og så er det et meget veldesignet studie, synes jeg, man kan tydeligt se når man læser studiet igennem, læser metodeafsnittet igennem, at der er blevet tænkt over alle de detaljer man skal tænke over før studiet er startet, der er en meget fin beskrivelse af eksklusionskriterier, der er en meget udførlig beskrivelse af de statistiske overvejelser med hensyn til størrelse af studiet og også med hensyn til hvordan man vil sikre sig en interrim undersøgelse og undervejs sikre sig patienternes sikkerhed, at tingene er i orden og at man kan stoppe studiet hvis det er nødvendigt. Og så forholder de sig egentlig meget nøgternt til tingene. Der er selvfølgelig også nogle ting man kan kritisere, det er der jo ved alle studier, men jeg synes overordnet set er det et godt studie og også et studie der viser, at hvis man arbejder sammen, det er jo et nationalt studie et Danmark hvor alle centre har deltaget, at hvis man arbejder sammen så kan man lave studier der er så store og vigtige at de også betyder noget rent praktisk og det er noget der har haft stor betydning for hvordan vi behandler hjertepatienterne i dag, i Danmark, og også har understøttet det internationalt. Så et godt studie som hænger sammen med det kliniske, og i hvert fald også rigtig veldesignet. Så det er jo egentlig to forskellige studier, det ene er metodologisk meget korrekt det andet er lidt mere hypotesegenererende studie som kræver at man skal kigge på lidt flere ting. Den er lidt mere afsluttende, den sætter et statement fast, det er altså den måde vi skal gøre det på rent klinisk, i praksis nr (2).

Question: *Det er jo meget interessant, og det har i hvert fald givet mig anledning til nogle opfølgende spørgsmål, men inden vi tager dem så vil jeg også gerne have at du prøvet at*

fortælle mig generelt lidt om hvad du tænker om begrebet forskningskvalitet.

Response: Jeg synes jo egentlig at forskningskvalitet er et meget bredt begreb, det afhænger jo meget af hvilken opgave man har man skal løse, man har jo både den kvalitative forskning og den kvantitative forskning, og jeg synes egentlig det er god kvalitet begge dele, hvis det er man har, hvis man har sørget for at tænke sig godt om i forvejen, og hvis man ligessom bruger de korrekte forskningsmetodologiske tilgange til tingene. Det er egentlig det jeg synes; god forskning det er hvis man har en systemiseret tilgang til en problemstilling. Hvad problemstillingen så er, det skal selvfølgelig være en relevant problemstilling, og sikre sig at det ikke er en problemstilling som en masse andre har undersøgt i forvejen, så skal det i hvert fald være fordi man mener de ikke har undersøgt det godt nok for at man ligessom også kan se fra starten af at det er noget som man kan, uanset om det viser det ene eller det andet, så giver det stadig en information der er relevant for læseren. Det er i det hele taget det, at man lige sætter sig ned når man laver et studie og gennemtænker resultaterne af studiet, hvilke muligheder kunne der være? Det synes jeg er en enormt god kvalitet ved et studie hvis man kan se at det er blevet gjort, for det gør alt andet lige, uanset om man får det ene eller det andet svar på et studie, et positivt eller negativt svar, så vil det stadig være noget som kan bruges for klinikerne, hvis det er fra lægens synspunkt, for forskeren generelt, indenfor andre fag end sundhedsvidenskab, det gælder jo helt generelt at det er noget der skal passe med et relevant emne og så skal det designes sådan at uanset om man får det ene svar eller det andet, så er det noget man kan bruge til noget. Så man kan jo sige rigtig meget om forskningskvalitet, man kan opgøre det på mange måder. Der findes jo nogle lidt sjove måder at opgøre det på synes jeg. Der er mange der gør meget i f.eks. impact factor og forskellige tidsskrifter og forskellige impact factors, og det synes jeg, der er vel også en eller anden sandhed i at hvis du f.eks. får en artikel i NEJM, som her artikel (1), så er der et vist filter og man ved godt at impact faktoren den ligger, jeg tror på en 25-30 stykker, så der er rigtig mange, der refererer til din artikel, så det er noget der sætter gang i nogle ting. Det er selvfølgelig rigtig fint, men jeg synes det er lidt forkert at bruge det som et decideret mål for forskningskvaliteten, fordi der også er nogle emner, hvis man har nogle måske lidt mere afgrænsede emner som kan være særdeles interessante og måske også give nogle perspektiver på sigt, de er måske ikke så relevante lige netop for den enkelte læge, derfor er det ikke noget der vil komme i et tidsskrift som NEJM.

Question: *Fordi det netop er generelt?*

Response: Det er sådan et generelt blad for alle læger i hele verden og på den baggrund, og det samme gælder jo Nature eller Science, sådan nogle lidt mere overordnede tidsskrifter som spænder meget bredt, jamen de ting man får deri det er noget som skal appellere til den generelle befolkning, så det er ikke nødvendigvis et tegn på forskningskvalitet, men man kan sige, hvis man får det i de tidsskrifter, de store internationale tidsskrifter, så er der jo en eller anden form for filter, det filter der er for at komme dertil, det er meget, meget kraftigt, altså det kommer igennem alle de forskningsmetodologiske

ting, det er virkelig gennemgået kraftigt. Statistikken skal jo være fuldstændig i orden før man får det deri, så det er et stempel for kvalitet når det kommer deri, og det er noget der er relevant. Men det er ikke det samme som at sige, at fordi det ikke kommer deri - ens forskning i de store tidsskrifter, med den høje impact factor - at det så er dårlig kvalitet. Fordi, hvis man nu har bare et lillebitte, nørdet område som man synes er spændende, og også andre kan se et perspektiv i, så kan det være at man kun har en impact factor, i det tidsskrift der hører til det emne, på 1-2 stykker, og så er det stadig noget der batter og har en god kvalitet. Så man kan opføre det på mange måder, men jeg synes det primære ligger i, at man er systematisk i sin tilgang til en problemstilling, og prøver at løse den ved falsifikation, altså helt almindelig grundlæggende Popper's falsifikation, som vi har lært i videnskabsteori i sin tid på studiet.

Question: *Nu nævner du selv impact factors, og du nævner det her med at de generelle tidsskrifter har en meget høj impact factor, men hvad nu hvis vi ser på et specifikt emneområde, det kunne være kardiologi f.eks., der har vi i hvert fald nogen tidsskrifter der, der også har en meget høj impact factor...*

Response: Det må man sige.

Question: *...og så er der nogle andre, der ikke er så høje...*

Response: Helt klart.

Question: *...er det så noget andet? Nu siger du, at en af grundene til at man ikke kan bruge impact factor, det er fordi de her generelle tidsskrifter har en meget høj impact factor, og det siger ikke nødvendigvis noget om relevansen. Men hvad nu hvis det er indenfor f.eks. et enkelt område, f.eks. Circulation, er der så en større sammenhæng imellem tidsskrifternes og artiklernes kvalitet i de tidsskrifter, og så impact factor, eller?*

Response: Jeg tror det er svært at sige, altså, man kan sige igen, de store kardiologiske tidsskrifter, Circulation, European Heart Journal, American Heart Journal, JACC, altså nogle af de der store amerikanske og europæiske tidsskrifter, jamen de er, det er jo god kvalitet det meste der kommer deri. Igen, der er et filter, fordi alle folk gerne vil have sine artikler i British Medical Journal, det er så et lidt mere generelt, men de kardiologiske der er et stort filter, og det gør alt andet lige, at kvaliteten bliver rigtig god, og det at der er en høj impact factor tyder også på at det sætter gang i nogle ting. Folk refererer til hinanden, og man kan sige det genererer ny forskning, så det er et kvalitetsmål, men det jeg måske synes er forkert er at det er det eneste kvalitetsmål. Det er jo en eller anden form for kvalitetsmål, synes jeg, det fortæller et eller andet om forskningskvaliteten, men der er jo bare også nogle mindre områder som ikke appellerer til den almindelige kliniker, som de her store kardiologiske tidsskrifter også refererer til, og det kunne være f.eks. nogle subting indenfor kardiologien, og hver gang man går længere ned i, tag Circulation f.eks., Circulation har jo lavet en masse sub-tidsskrifter under Circulation,

som har lavere impact factor end det regulære tidsskrift, og det er jo fordi ligeså snart man begynder at gå ned i små detaljer, så ryger impact factoren også stille og roligt ned, fordi man kommer ind til et mindre og mindre publikum med et mindre og mindre behov for at citere hinanden hele tiden. Så jeg tror egentlig der gælder lidt det samme indenfor kardiologien som der gør for de generelle. Men det er også fordi de store kardiologiske tidsskrifter de er relativt generelle, Circulation er et ret generelt tidsskrift for kardiologer og kardiologi fylder i det hele taget enormt meget for interne medicinere. Det fylder meget for læger, altså en enormt stor del af de kontakter der er i sundhedsvæsenet det er jo noget der er relateret til kardiologi, både ved akutte patienter og ved kroniske patienter. Så det er noget der har stort impact, ikke for at bruge impact factor, men generelt, ting indenfor kardiologi. Så jeg tror der gælder lidt af de samme ting som der gør for NEJM og Science og Nature osv. i den sammenhæng.

Question: *Er der andet som du - altså impact factor er jo afledt af citationer - er der andet som du kunne forestille dig man kunne gå ind og se på for objektivt at vurdere kvalitet?*

Response: Når man sidder og læser en artikel så går man den jo kritisk igennem, så ser man efter om der er nogle kvalitetsting der er opfyldt, men tænker du på mål som sådan? Tænker du talmål?

Question: *For eksempel*

Response: Det ved jeg ikke, nej...

Question: *En måde hvorpå man kunne kvantificere, og sige at den ene artikel er bedre end den anden, f.eks.?*

Response: Altså vi bruger jo i hvert fald indenfor klinisk medicin eller indenfor kardiologi, der bruger vi jo meget det her med at gå ind og vurdere typen af studie, og ligesom sige "hvor høj er kvaliteten på en skala fra, altså, er det A, B, eller C-rekommendationen f.eks. for en guideline, alt efter hvor meget information man har om et emne. Er det bare en kasuistik om et eller andet emne eller er det en parret, randomiseret studie, med et højt antal som kan være med til at understøtte en problemstilling og fortælle noget om en problemstilling. Så selve designet af studiet er i sig selv et mål som man bruger når man vil se på, i hvert fald hvilken betydning resultaterne har efterfølgende for hvordan vi behandler vores patienter. Det er noget man bruger meget når man skal ind og lave rekommandationer, men jeg tror ikke jeg ved helt hvad du fisker efter?

Question: *Jeg fisker ikke efter noget bestemt, jeg vil egentlig gerne bare høre hvad du har. Det du nævner her nu, det er sådan noget med evidens?*

Response: Det er evidensniveauer, ja, men det er så også ofta at få flere studier poollet sammen, om en eller anden problemstilling. Ellers så synes jeg meget af kvaliteten går på hvad for et design man har valgt. Designet skal jo også være rigtigt til formålet. Man

kan jo sagtens lave en... nogle ting lader sig jo ikke undersøge i randomiserede studier, sådan er det jo bare, så må man nøjes med observationelle studier i den sammenhæng, og det kan jo også være ganske fint hvis de er designet på den rigtige vis og man har tænkt sig godt om før man får sine resultater. Men på en eller måde så er det jo sådan at mange vil sige at randomiserede studier, det er jo ligesom kongen af studier.

Question: *Vil du sige at der måske er en højere klinisk værdi jo højere evidens der er? Giver det mening at sige det? Hvis man har to studier, hvor det ene er et dobbelt-blindet randomiseret forsøg og det andet det er en case-serie, vil du så sige at det dobbelt-blindede, randomiserede forsøg har en højere værdi for f.eks. en kliniker eller en forsker?*

Response: Jamen det har det jo, hvis det er sådan at det randomiserede studie besvarer det spørgsmål, der er blevet stillet. Så har det en større værdi. Hvor man kan sige - igen, man skal passe på, for den case-serie man nu har, som du nævner som andet alternativ, hvis nu den case-serie rent faktisk peger nogle ting frem, som er hypotesegenererende for andre, der kan sætte nogle ting igang, jamen så er det jo også god kvalitet, og man må alt andet lige formode, at hvis folk vælger at præsentere det som en case i et større tidsskrift, eller i et tidsskrift, så er det fordi man mener det har et eller andet, som kan bruges af andre til at påpege en pointe eller alternativt sætte et studie igang fordi der er noget, der mangler at blive undersøgt her i den her patientgruppe. Så jeg synes igen man skal passe på med at sige, at det ene eller det andet er kvalitet. Man må sige at det er... jeg synes det afhænger meget af om der er det rigtige spørgsmål og om der er det rigtige design til det spørgsmål, og det kan man afgøre om det er en kvalitet ud fra om man har tænkt sig godt om i forvejen. En kasuistik kan man jo diskutere hvad det er, altså hvor meget forskning der er i en kasuistik, men hvis man har nogle case-serier, eller har noget observationelt, hvor man laver nogle gode betragtninger, lidt à la den her, jamen det er jo egentlig bare en deskriptiv opgørelse over nogle ting, men hvis den kan sætte nogle andre ting igang, jamen så synes jeg, så er det god kvalitet. Og hvis man har tænkt over den måde man gør det på, hvis man har gjort det elegant, hvor man kan sige et randomiseret studie som det her, det er også rigtig godt, designet godt til at besvare det spørgsmål der er. Det ville jo også være dumt hvis det var sådan at man havde lavet et stort tungt, dyrt randomiseret studie hvis man kunne besvare spørgsmålet ved noget meget mere simpelt, observationelt. Så det elegante, den gode forskning ligger jo også i at man vælger sin metode efter sit problem og lægger sin arbejdsindsats ind derefter. Det ville jo være fuldstændig åndssvagt at bruge tonsvis af kroner og energi på noget, hvis man kunne i hvert fald give et indtryk af at der er noget at komme efter via en mere simpel, observationel metode.

Question: *Du nævnte på et tidspunkt i starten, da du fortalte om de her artikler, at ved den ene var der et samarbejde med industrien. Eller man benyttede sig af en database... vil du mene at sådan noget med at der er et samarbejde, det kunne både være imellem den enkelte forsker, på f.eks. et sygehus, og så medicinalindustrien eller medikoteknisk industri, eller det kunne være samarbejde imellem flere forskere på flere sygehuse, har det noget at sige*

i forhold til kvaliteten af det der kommer ud, eller kan en enkelt forsker lige så vel skrive høj-kvalitetsstudier som forskningsgrupper kan?

Response: Jamen altså en enkelt forsker kan selvfølgelig godt skrive høj kvalitetsstudier, det er der jo talrige eksempler på, men man må jo sige at der er mange ting som ikke er mulige hvis man sidder som enkeltforsker, og jeg tror også at ved at man får flere personer involveret, så får man også et bedre produkt. Altså begge de artikler jeg har taget med, det er jo artikler som, især det randomiserede studie her, det er jo en artikel som har mange medforfattere, som har mange øjne med inde over, og man kan helt tydeligt se, at det ikke er en kliniker der har siddet og lavet statistikken, det er helt klart en statistiker der har været inde over og lave en rigtig god beskrivelse. Der er jo også mange eksempler på folk der har siddet alene med studier, og sidder og får deres resultater tilbage, og så kan man se at hvis der nu havde været nogle andre folk med inde over fra starten så kunne de se at den del her, den forskningsmetodologiske process, så havde man fået et meget bedre produkt. Så jeg er stor tilhænger af, at man laver forskning i grupper frem for at sige kun at lave det selv. Altså det giver alt andet lige nogle bedre resultater, man får jo også nogle idéer når man sidder i grupper, det er jo en god kogegryde til at få nye idéer, og nye studier, når man sidder flere personer sammen, jeg synes det er en stor styrke, og ofte også en kvalitet, man tror også mere på studier - altså jeg har det sådan, når jeg sidder og læser et studie, bare fordi det er publiceret, så skal man alligevel læse det kritisk, og det er klart at hvis der står flere forfattere, så jeg ved hvad står for, som jeg har set og snakket med og læst andre studier, som jeg ved hvilken kvalitet har, det i sig selv kan også være med til at løfte min opfattelse af kvaliteten af studiet. Men det at man ved det er flere forskellige baggrunde de her forfattere kommer med, så får man også et produkt som er mere gennemtænkt, hvor man ved at så mange ting som muligt er blevet overvejet i processen. Så jeg mener bestemt at det er noget som er med til at løfte kvaliteten, som udgangspunkt, der er sikkert også eksempler på det modsatte, for man kan også få for mange input.

Question: *Nu har vi snakket meget om forskningspublikationer, vi har også snakket lidt om at der også ligger noget forskning bagved. Er det sådan at der er en direkte sammenhæng imellem forskningen og forskningspublikationerne i forhold til kvalitet?*

Response: Hvad tænker du der?

Question: *Er det sådan at en god artikel altid er resultatet af et godt studie, eller, for at man kan skrive en god artikel, kræver det så at der ligger et godt studie bag? Eller kan man godt skrive en god artikel på baggrund af et dårligt studie*

Response: Altså, der er jo en vis sammenhæng. Hvis du har et godt materiale, så får du formentlig også et bedre produkt, hvis du ser det ud fra din artikel efterfølgende. Men hvis man er i stand til at forholde sig kritisk til den forskning der er lavet når man laver sin artikel, så kan produktet jo stadig være ganske udmærket. Det vil jo være

sådan i mange sammenhænge, så når folk de sætter et studie op, så når de får deres resultater ind, så viser det sig, at det ikke var helt som forventet, og så må man jo i sine limitations fortælle hvad man mener problemstillingerne er ved artiklen. Men derfor kan det jo stadig have en værdi, det kan stadig være god forskningsværdi, selvom der er nogle problematiske ting omkring et studie.

Question: *Det lyder på dig som om det godt kan have en værdi, men dog ikke lige så høj som hvis det var et godt studie?*

Response: Altså som udgangspunkt så må man sige at det er - altså, jeg synes det er ligesom et korthus, hvis man har et godt fundament, så har man også et bedre udgangspunkt for at man kan komme med en bedre top, hvor man kan sige at publikationen er toppen på forskningen, for det er jo det, der bliver videregivet til de andre forskere der arbejder med det rundt omkring i verden. Hvis det er sådan at man kun har alt til at ligge nede i bunden af en skuffe, så har det ikke nogen relevans alligevel. Så det der bliver videregivet har stor betydning, men hvis det er sådan at fundamentet ikke er godt, så vil produktet i mange sammenhænge heller ikke blive godt. Men hvis man har et dårligt fundament, og man er i stand til at formidle det dårlige fundament, hvis man er godt til at skrive og nøgternt fremstille de problemstillinger der var når man lavede sit studie, så kan man stadig godt få en god artikel ud af det, fordi man tager højde for at der var nogle problemstillinger undervejs, der ikke var i orden. Og så længe folk de er ærlige omkring hvad de skriver, så synes jeg, så kan man sagtens få en ganske udmærket artikel ud af det. Det er klart at et bedre fundament giver et bedre produkt. Man kan også godt ødelægge en artikel, den anden vej rundt, hvis man ikke skærer fuldstændig ind til kernen, f.eks. det her studie her omkring ballonbehandling, det er jo et godt eksempel på et studie, hvor der er et kæmpe setup, der har været mange med til at lave det og der var rigtig mange data der blev sendt ind, men det man har valgt at gøre når man laver publikationen, det er man har valgt at fokusere på ganske få ting, og det der egentlig er præsenteret af data deri det er ganske lidt. Så man har valgt at fokusere på det der er klinisk relevant, og alt andet støj har man ikke taget med i artikler. Og det eksisterer, jeg ved at det eksisterer, men det der er med i artiklen, det er det, der er relevant for klinikerne, så det er let læseligt, den besvarer et præcist spørgsmål man kan arbejde videre med. Der er også noget i den proces med at være i stand til at kunne koge tingene ind til en lille pointe til sidst, som er det man skal tage med som budskab. Hvis man ud fra det samme materiale havde skrevet en artikel som var ti gange så lang, hvor der havde stået tyve pointer i, og det kunne man godt have gjort, så mener jeg det ville være en dårlig artikel, for så ville den ikke få den samme impact klinisk, for folk ville ikke læse artiklen, folk ville ikke forholde sig til den på samme vis, fordi det ville være at overgøre det. Selvom det er det samme materiale, der bliver brugt.

Question: *Så ville du hellere, at man delte det op i flere artikler?*

Response: Så vil jeg hellere have man deler det op i flere artikler, hvis det er sådan, at

det er relevant i forhold til det kliniske budskab, hvis det er relevant, hvis det gør at man får sit budskab bedre ud. Jeg synes man skal passe på med at lave artikler ud fra salami-metoden, hvor man stille og roligt sprøjter en masse artikler ud fra et enkelt studie, men nogle gange kan det være fornuftigt nok at adskille tingene, for det vil jo være sådan at man skal sætte sine resultater op imod de givne studier der er indenfor et emne og det bliver jo en frygtelig rodet diskussion hvis det er sådan at man skulle have ti forskellige problemstillinger dækket i én artikel, det ville ikke være rimeligt overfor læseren, og det ville ikke være rimelig overfor noget som helst, så er det ingenting værd, så er det bedre at dele det op i nogle lidt mere afgrænsede problemstillinger og så forholde sig kritisk i forhold til den eksisterende metode der er.

Question: *Helt til sidst har jeg en lille brainstorm-session, der er nogle ord jeg vil nævne, og så vil jeg gerne have du bare siger det første du tænker på. Det må være kort, det må være langt, det bestemmer du selv. Det første ord jeg har, det er ny viden*

Response: Ny viden... nu sidder jeg og tænker på hvorfor i lige den her sammenhæng. Det er jo altid godt når et studie bidrager med ny viden, det er jo en af pointerne med at lave forskning, at man bidrager med ny viden, der ikke eksisterede. Jeg ved ikke hvad jeg skal sige ellers... ny viden er vigtig i forskning, selvfølgelig er det det

Question: *Tværfaglighed?*

Response: Tværfaglighed det er en absolut nødvendighed for at lave god forskning i de fleste sammenhænge, og det er noget der har løftet kvaliteten af forskningen i Danmark, og også internationalt tror jeg, at man er blevet bedre til tværfagligt samarbejde, og benytte sig af at der er nogen, f.eks. statistikere, som har en indsigt i noget der er meget komplekst, at man kan benytte sig af at få dem ind meget tidligt i processen og deltage aktivt som en del allerede i designfasen, frem for at de først bliver præsenteret for tingene bagefter, så det at man får folk ind, får det tværfaglige ind, udnytter at de har nogle forskellige ting, det øger klart kvaliteten i mange ting.

Question: *Er det også noget du ser? Altså, er vi gode til at være tværfaglige, eller er det noget vi er på vej til?*

Response: Det synes jeg. Det er en proces der har været i gang over flere år, og mit indtryk er at det nok er over de sidste ti år, hvor der er sket en betydelig udvikling i tværfagligheden indenfor forskningen, man kan også se at der sker en eksplosiv vækst i antallet af medforfattere på diverse artikler, det kan jo både være godt og skidt, men mit indtryk det er at man er begyndt mere at anerkende, at folk har nogle forskellige ting at byde ind med, og det giver anledning til bedre produkter. Nu ligger det også lidt i tidsånden set ud fra den måde man har tilrettelagt undervisningen på universiteter, og andre videregående uddannelser, at man lægger mere op til at tingene skal være i gruppearbejde. Her i Aalborg har man jo også en tradition, ude fra Aalborg Universitet,

at man ser gruppearbejde som et gode i forbindelse med at udvikle ting. Så det er også det man bruger i industrien, og jeg synes det er en positiv ting, tværfaglighed, hvis man bruger det rigtigt.

Question: *Impact?*

Response: Impact kan jo være mange ting, nu har vi lige siddet og snakket om impact factor, jeg synes egentlig ikke det er det første jeg tænker på, jeg tænker mere på hvilken betydning det har klinisk, jeg har en læges øjne på når jeg ser på forskning, så jeg ser ofte efter om det er noget der har betydning for vores kliniske dagligdag, så det at et studie f.eks. har en god klinisk impact er noget der betyder noget for mig. I det hele taget betyder impact jo bare at ens studie batter noget, på den måde at det kan sætte noget andet igang, kan bidrage med noget ny viden og ligesom giver en bedre beskrivelse af den problemstilling der forlår fra starten. Impact er jo et kvalitetsmål i sig selv, det er bare hvordan man opgør det, om det er en factor eller om man giver en beskrivelse af det i stedet for, altså mere præciseret, en sproglig beskrivelse i stedet for bare et tal.

Question: *Citationer?*

Response: Citationer betyder jo noget i mange sammenhænge, det at noget bliver citeret meget ligger jo også i impact factor. Det fortæller noget om at det er noget, der interesserer folk. Man har enten provokeret folk med et eller andet, det kan også være folk er enige, men ofte er folk uenige, ofte er det en provokation, det kan også være man har lavet et rigtig godt studie, og folk gerne vil kommentere på det, eller at det har sat nogle andre ting igang. Det at ting bliver citeret meget, kommenteret, men også citeret i andre studier tyder jo på at andre forskere har fundet det materiale man nu har publiceret brugbart i deres arbejde og dermed har det også bidraget til udviklingen. Hvis man har et studie som kommer ud og aldrig nogensinde bliver citeret i andre sammenhænge, så kan det godt være at man har været med til at klarlægge en problemstilling, men det har så ikke været en problemstilling som har sat andre folk igang med andre ting, så det er måske ikke noget der har bidraget så meget til udviklingen. Så det fortæller jo noget om bidragets udvikling i en eller anden grad i hvert fald.

Question: *Har du helt tilfældigvis været inde og se hvor citeret de her to artikler er?*

Response: Nej, det har jeg ikke.

Question: *Du har ingen fornemmelse af det?*

Response: Jo, jeg vil umiddelbart tro at, eller jeg ved at det her er ekseptionelt, altså nr. (2) her, den er blevet citeret rigtig mange gange, det er jeg helt overbevist om, altså vi snakker flere hundrede gange, det tror jeg. Det er også et af de tunge studier, i hvert fald indenfor dansk kardiologi. Det andet her (1), det ved jeg faktisk ikke, hvor meget det er citeret. Det har jeg ikke været inde og undersøge, fordi for mit vedkommende,

til det jeg har skullet bruge det til, da har det ikke været, det er ikke noget jeg har brugt til at måle kvaliteten. Som sagt synes jeg man skal passe på med at måle kvaliteten ud fra et tal. Det fortæller selvfølgelig noget om at, hvis det er blevet citeret mange gange, at det er noget der har interesse for mange mennesker, noget folk kan bruge til noget. Man kan godt bruge det til noget i en vis grad, men ikke noget entydigt.

Question: *Layout af artikler*

Response: Det jeg lige umiddelbart tænker er, at hvis man vil have sit budskab frem i en artikel, så er det vigtigt man får lavet et præcist og kort præsentation af den problemstilling der er, og med hensyn til layout, om det er i farver eller hvad det er tror jeg egentlig ikke er så vigtigt. Ofte så er abstract og titel det der fanger ens opmærksomhed, og det er det der betyder noget, at man har sat det op på en god måde, så går folk videre til at læse noget mere. Men det er vigtigt at man opretholder den stringens som man i øvrigt bruger i forskningsartikler når man laver sit layout, med IMRAD. Man skal selvfølgelig overholde de konventioner der er hvis man vil have andre forskere til at tage en seriøst. Hvis man ikke overholder konventionerne så vil man allerede der begynde at stille spørgsmål tegn ved kvaliteten. Det gør at det er meget lettere at læse et studie hvis man ved det er opstillet efter de korrekte, eller den gængse måde at opstille en artikel på og man ved at folk forholder sig objektivt i en del og måske mere subjektivt i en anden del, hvor man har nogle meninger inde over også, så man ligesom ved hvad der hører til hvad.

Question: *Det sidste jeg har det er peer review*

Response: Det er jo en vigtig ting. Det at man får kolleger som man ikke kender til at sidde og læse sit studie igennem og komme med kommentarer til studiet det vil til enhver tid øge kvaliteten af et studie. Hvis man har et studie der er publiceret i et tidsskrift hvor der ikke er peer review på, det vil jeg været bekymret for, det vil være noget der vil skærpe min kritiske sans betydeligt. Jeg synes peer review er en absolut nødvendighed for at man får bedre publikationer, bedre forskning i det hele taget, nu har jeg selv været involveret i flere, og lavet flere peer review og jeg synes man får altid noget ud af det, både som den der laver peer reviewet, men som modtager får man også rigtig meget ud af det, hvis man ser lidt nøgternt på de kommentarer man får. Der er stort set altid substans i kommentarerne og slutproduktet, selvom det måske tager noget tid ekstra at lave, bliver altid bedre. Det er bestemt en vigtig ting i forskningskvalitet at man har peer review inde over.

References, pilot study

Andersen, H. R., Nielsen, T. T., Rasmussen, K., Thuesen, L., Kelbaek, H., Thayssen, P., ... DANAMI-2 Investigators (2003). A comparison of coronary angioplasty with fibrinolytic therapy in acute myocardial infarction. *New England Journal of Medicine*, 349(8), 733-742.

- Bokhour, B. G., Clark, J. A., Inui, T. S., Silliman, R. A., & Talcott, J. A. (2001). Sexuality after treatment for early prostate cancer. *Journal of General and Internal Medicine*, 16, 649-655.
- Bondas, T. (2009). Preparing the air for nursing care: A grounded theory study of first line nurse managers. *Journal of Research in Nursing*, 14(4), 351-362.
- Brahe, L., & Sørensen, E. E. (2010). Afrydelser i klinisk sygeplejepraksis. *Klinisk sygepleje*, 24(1), 29-41.
- Cannon, C. P. (2004). Intensive versus moderate lipid lowering with statins after acute coronary syndromes. *New England Journal of Medicine*, 350(15), 1495-1504.
- Durbin, J., & Koopman, S. J. (1997). Monte carlo maximum likelihood estimation for non-gaussian state space models. *Biometrika*, 84(3), 669-684.
- Dwyer, J. H., Allayee, H., Dwyer, K. M., Fan, J., Wu, H., Mar, R., ... Mehrabian, M. (2004). Arachidonate 5-lipoxygenase promoter genotype, dietary arachidonic acid, and atherosclerosis. *New England Journal of Medicine*, 350(1), 29-37.
- Foss, C., & Ellefsen, B. (2002). The value of combining qualitative and quantitative approaches in nursing research by means of method triangulation. *Journal of Advanced Nursing*, 40(2), 424-248.
- Harrison, P. J., & Stevens, C. F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(3), 205-247.
- Hauser, R. G., Katsiyannis, W. T., Gornick, C. G., Almquist, A. K., & Kallinen, L. M. (2010). Deaths and cardiovascular injuries due to device-assisted implantable cardioverter-defibrillator and pacemaker lead extraction. *Europace*, 12, 395-401.
- Joensen, A. M., Overvad, K., Dethlefsen, C., Johnsen, S. P., Tjønneland, A., Rasmussen, L. H., & Schmidt, E. B. (2011). Marine n-3 polyunsaturated fatty acids in adipose tissue and the risk of acute coronary syndrome. *Circulation*, 124, 1232-1238.
- Llewellyn, S. (2001). Two-way windows: Clinicians as medical managers. *Organization Studies*, 22(4), 593-623.
- Mozaffarian, D., Psaty, B. M., Rimm, E. B., Lemaitre, R. N., Burke, G. L., Lyles, M. F., ... Siscovick, D. S. (2004). Fish intake and risk of atrial fibrillation. *Circulation*, 110, 368-373.

A.2. Main study transcriptions

A.2.1. Interview reference: M01

Date: Oct 12, 2011

Time: 14:00

Question: *Jeg har bedt dig tage de her artikler med, på baggrund af at du har vurderet dem som høj kvalitet. Det vil jeg meget gerne høre dig fortælle lidt om, hvad det er du vurderer som høj kvalitet?*

Response: Jeg ved egentlig ikke om det er højere kvalitet end alle mulige andre artikler jeg har læst, jeg har taget dem ud fra min kliniske dagligdag. Det er nogle artikler jeg har brugt i min dagligdag, netop lige præcis det her emne omkring HER2 diagnostik ved gastrisk cancer, det er vi ved at indkøre på instituttet, og igen, det her er også en artikel om den kliniske dagligdag, når der opstår et problem omkring diagnostik af en bestemt sygdom, så har jeg brugt den her. Så jeg ved ikke hvad der gør den af høj kvalitet, men

i og med jeg har brugt den i min kliniske dagligdag synes jeg den har en høj kvalitet.

Question: *Det er simpelthen anvendeligheden der gør det?*

Response: Lige præcis, anvendeligheden.

Question: *Er der nogle punkter hvor den skiller sig ud fra andre artikler? Altså, kunne du ligeså godt have fundet en anden artikel om det samme emne, som havde været ligeså god?*

Response: Sikkert ikke, fordi netop førsteforfatteren og en af de andre forfattere er sådan nogle hotshots indenfor det her område, og indenfor patologien. Og det her er ligesom referenceartiklen for alle mulige andre artikler der er lavet efterfølgende, den her er ikke så gammel, den er fra 2010, snart to år gammel. Men det er ligesom dem der har lavet en stor undersøgelse af et kæmpestort materiale så af den grund gør det den jo også til høj kvalitet.

Question: *Okay, så de har meget stort datamateriale, og de er nogle af de første også?*

Response: Ja, jeg tror de har omkring 550-600 patienter i den her første indledende undersøgelse, hvor man har fundet ud af at man kan bruge et bestemt stof i den onkologiske behandling af de her patienter. Og så er der ligesom kommet flere studier ovenpå det her. Så de har ligesom været banebrydende.

Question: *Den anden artikel, kan du uddybe lidt på den?*

Response: Ja, det er jo så fordi jeg er stødt på et problem i min kliniske hverdag angående i forbindelse med diagnostik af spiserørsbiopsier, hvor man har snakket meget om en bestemt tilstand der hedder eosinofil esofagik, og der står ikke så meget i lærebøgerne, så der har jeg tyet til artikler for at se hvordan man diagnosticerer de forskellige ting. Netop fordi den måske ikke er hyppigt forekommende og den er ikke så beskrevet.

Question: *Nu nåede jeg ikke lige at se det her... din forskningserfaring, du har skrevet "nogen", men det er tidligere...*

Response: Ja, jeg har ikke nogen... vi har jo enormt travlt i vores kliniske hverdag, så det er ret begrænset hvad det bliver til af...

Question: *Men det er jo interessant for mig lige at kunne skelne imellem hvem forsker nu, og hvordan forholder de sig, og hvem forsker ikke... så det er jo en anden indgangsvinkel du har til det end en der var aktiv forsker?*

Response: Ja

Question: *Ja, og det er simpelthen det kliniske i det?*

Response: Lige præcis

Question: *Og det er jo også rigtig interessant... men inden jeg følger op på det, så vil jeg gerne lige spørge; når jeg bare siger forskningskvalitet, som begreb, hvad tænker du så om det, hvis vi ser bort fra de her to artikler, men i stedet ser mere bredt?*

Response: Så er det noget med et stort materiale, hvis det er derhenad du mener?

Question: *Det kan det sagtens være, ja...*

Response: Og så at det er nogen der har en vis forskningserfaring også, dem der har lavet den forskning. Det behøver det jo heller ikke at betyde, det er jo egentlig lidt forkert at sætte det sådan op, fordi folk der er igang med deres uddannelse til speciallæge kan jo også lave fantastiske artikler alligevel, jaeh, det ved jeg ikke.

Question: *Altså du må sige lige hvad du vil, der er ikke noget der er forkert.*

Response: Nej, det ved jeg faktisk ikke.

Question: *Hvis nu... det kan godt være det bliver svært at svare på så, men hvis nu man skulle prøve objektivt at evaluere kvaliteten af noget forskning; først og fremmest, kan man det? Og er der noget du ville gå ind og se på?*

Response: Det tror jeg ikke jeg har forskningserfaring nok til at kunne fortælle om. Så det ved jeg faktisk ikke.

Question: *Det er bare helt i orden. Så vil jeg prøve at nævne nogle ting, som måske kunne være muligheder. Sådan noget som evidensniveau, f.eks., er det noget du tænker over, når du læser en forskningsartikel? Sådan noget som at det er dobbeltblindet, randomiseret forsøg, kontra et case-studie.*

Response: Jamen det er klart, det gør det jo af en højere værdi, at det er sådan nogle dobbeltblindede forsøg. Men det findes der ikke så meget af indenfor vores verden. Det er mere sådan nogle opfølgninger eller retrospektive undersøgelser af et stort patientmateriale der ligger i forvejen. Sådan er det i hvert fald på meget patologi.

Question: *Er det så derfor du vurderer størrelsen som meget vigtig?*

Response: Lige præcis, ja.

Question: *Det kan måske ligefrem være at patologien skiller sig lidt ud fra de andre medicinske specialer?*

Response: Nej, for man kan jo godt lave forsøg og undersøgelser, der er dobbeltblindede og så videre, som har de der høje kvaliteter og så videre, men det tager bare rigtig

lang tid at få fundet de patienter som man skal have. Og vi har jo et stort arkivmateriale indenfor patologien, så der ligger jo masser af cancer og biopsier og DNA som man kan bruge.

Question: *Men hvis man nu tænker det her med evidens lidt nærmere, så hører jeg nogen sige at det er selve designet der er dét, der giver en kvalitet. Altså dét at det er et dobbeltblindet forsøg er per definition kvalitet i sig selv, hvorimod andre siger mere, at det er sammenhængen imellem designet og problemstillingen. Hvordan vil du vurdere det?*

Response: Det ved jeg faktisk ikke.

Question: *Helt i orden. Så hører jeg også nogen sige, at der er en slags fødekæde hvor man kan sige, at det godt kan være det randomiserede forsøg er det vi stræber imod, men man kan ikke lave det medmindre der ligger andet før. Ser du det også sådan? Giver den metafor med en fødekæde mening?*

Response: Ja, der skal være noget før man selv kan gå igang, det er ligesom det jeg hører dig sige.

Question: *Man kunne også oversætte det til, at der skal være noget grundforskning inden der kommer noget anvendt.*

Response: Jamen selvfølgelig skal der det, det er helt nødvendigt. Det er derfor man er så afhængig af grundforskningen, også indenfor lægevidenskaben. Og især indenfor patologien. Så jo, det giver da meget mening. Også det med idéerne; det er jo ikke alle der får de der forkromede idéer, så mange finder vel inspiration i andre forskningsprojekter.

Question: *Når man så arbejder med patologi, hvad så med kliniske retningslinier, er det stadigvæk interessant for jer?*

Response: Ja, i nogle sammenhænge er det da. Vi har også vores egne retningslinier inde på instituttet, i form af PRI-dokumenter, som også er en klinisk vejledning, med udskæring af forskellige præparater, tarmpræparater og sådan noget. Så jo, det har da en vis betydning.

Question: *Og det... bliver de udarbejdet på samme måde som de øvrige?*

Response: Ja, som kliniske retningslinier? Ja.

Question: *Så det kunne man også sige er det sidste led i fødekæden?*

Response: Ja, det er det der er evidensniveau Å eller hvad nu...

Question: *Så vil jeg lige høre om du helt tilfældig har været inde og se hvor ofte de her artikler er citeret?*

Response: Nej, det har jeg faktisk ikke. Jeg kunne forestille mig at sådan en som den her var ret meget citeret, men nej.

Question: *Så har jeg nogle begreber som jeg gerne vil høre din umiddelbare reaktion på. Selvfølgelig i konteksten forskningskvalitet, og du bestemmer selv om du vil sige noget langt eller kort, om du er positiv eller negativ i det du siger, det er bare din første indskydelse. Det første begreb er ny viden.*

Response: Så skal jeg bare sige min umiddelbare? Jamen ny viden får man fordi man forsker. Så derfor er forskning nødvendig for at man kan få ny viden. Jeg ved ikke hvad jeg mere skal sige om det.

Question: *Tværfaglighed?*

Response: Det er for det meste godt, men kan også være dårligt. Men jeg ser helt sikkert tværfaglighed som en styrke indenfor forskning. Man bliver nødt til at have sine klinikere... altså jeg, som patolog, bliver nødt til at have mine klinikere med hvis jeg skal lave noget forskning - noget fremadrettet forskning i hvert fald. Så jo...

Question: *Så I er tvunget til at samarbejde?*

Response: Ja, det er vi.

Question: *Hvis der så er - jeg ved ikke om det giver mening at sige det - en patologisk grundforskning; så er tværfaglighed måske knap så vigtig?*

Response: Jamen så kan du gå hen i dit arkiv og finde noget, hvis du undersøger en bestemt tumors ekspression af et eller andet, så behøver man jo ikke bruge sine klinikere, der kan man bruge sit arkiv.

Question: *Hvad med sådan noget som impact factor? Siger det dig noget i den sammenhæng?*

Response: Det er jo noget med citering, eller hvordan?

Question: *Ja, det har noget med det at gøre. Det er tidsskriftet specifikt; hvor ofte artikler i tidsskriftet i gennemsnit bliver citeret.*

Response: Nej, det har jeg ikke nogen mening om... men det er jo klart at jo højere en impact factor en artikel, tidsskrift har, jo mere anvendt er ens forskning måske, ikke at den nødvendigvis er bedre.

Question: *Det kunne også være et udtryk for andre ting?*

Response: Ja, at man skulle have at vide hvordan man ikke skulle lave noget.

Question: *Ja, det kunne det selvfølgelig også være... Det næste jeg har er citationer direkte til den enkelte artikel. Hvad tænker du om det?*

Response: Det ved jeg ikke, men jeg kunne forestille mig at folk der var banebrydende indenfor et område de var citeret rigtig mange gange, og det var nok efterhånden mere af navn end af gavn, kunne man forestille sig i nogle tilfælde. Selvfølgelig er det også fordi dem der bliver citeret mange gange også laver den mest banebrydende forskning og har gjort det i lang tid, og har ressourcer til at lave det. Men ellers så er det jo dem der bliver citeret flest gange der er hotshots.

Question: *Layout af artikler?*

Response: Det synes jeg betyder utrolig meget... jeg synes de her ser røvedelige ud.

Question: *Så med layout da tænker du opsætning, rent grafisk?*

Response: Ja.

Question: *Som det sidste punkt har jeg så peer review.*

Response: Det siger mig ikke så meget. Andet end at artiklerne kan være peer reviewed.

Question: *Men du har ikke lige nogen fornemmelse af, om det kan være vigtigt?*

Response: Det er jo selvfølgelig vigtigt, det er vel vigtigt for om en artikel kan blive optaget.

Question: *Men er det vigtigt at vi har det?*

Response: Det er det vel, det ved jeg ikke, jeg kender ikke så meget til det, at jeg kan udtale mig om det.

Question: *Hvor meget bruger du forskningsartikler i dit arbejde, altså læser dem, er det noget du gør ugentligt, månedligt?*

Response: Det gør jeg ugentligt, jeg har ikke lige noget tal på hvor mange jeg får læst, men det er ugentligt jeg bladrer artikler igennem. Netop i sådanne tilfælde her, hvor man lige søger noget eller lige studser over noget.

Question: *Så det er både aktiv søgning efter det og så er det opdatering?*

Response: Ja, det er det.

Question: *Når du så vælger een artikel ud frem for en anden, er det så på baggrund af en eller anden relevans?*

Response: Det er klart, men også på baggrund af hvor nye de er, det synes jeg også har betydning. Jo nyere de er, jo mere opdaterede.

Question: *Men det der så er afgørende for om en artikel den er relevant for dig, kan du fortælle mig en lille smule om det?*

Response: Jamen det er jo om det er et emne jeg måske beskæftiger mig med i min kliniske dagligdag. Så bliver den utrolig vigtig, fordi den så danner grundlag for hele min kliniske dagligdag.

Question: *Så det er det tematiske?*

Response: Ja, det er det.

Question: *Vil du sige at indenfor det du specifikt arbejder med, vil du da sige der bliver skrevet så mange artikler at du ikke kan følge med i dem allesammen, eller er det så relativt få, at du godt kan overskue at læse dem allesammen?*

Response: Nej, der er alt for mange til at man kan nå at følge med i det. Så det er derfor jeg vælger ud efter de der kriterier med hvor nye de er i forhold til relevansen af dem.

Question: *Jamen så har jeg egentlig ikke mere. Medmindre du sidder med noget du mangler at fortælle noget om, eller du har noget til forskningskvalitet helt generelt?*

Response: Nej, det har jeg faktisk ikke.

Question: *Det er bare helt i orden. Så siger jeg mange tak.*

A.2.2. Interview reference: M03

Date: Nov 08, 2011

Time: 14:30

Question: *Vi har fundet de her to artikler, og nu vil jeg meget gerne høre hvad der gør...*

Response: ...hvad der gør at de står ud som noget særligt?

Question: *Netop.*

Response: Reference (1); det der helt klart tænder den som noget særligt det er ikke egentlig design, det er idé. Det er fordi idéen er ny, altså grundlæggende er metoden i orden, den er ikke ekceptionel, men den er i orden. Det er en sammenlignende undersøgelse, men idéen bagved er ikke prøvet før, og resultatet har så vist sig i de 20 år

der er gået siden, at det her har forandret verden, eller der er kommet en række andre undersøgelser som ligesom har benyttet sig af det design, og har dokumenteret det den artikel så fandt som det første. Altså genundersøgt og genfundet. Idéen var ny. Og det synes jeg ligesom får den til at stå ud som noget særligt.

Question: *Så der var simpelthen en ny idé?*

Response: En idé, ja. Det andet studie; kvaliteterne ved det er faktisk... der er flere ting der gør, at det også står ud som noget særligt. Den ene ting er at det er, det er selvfølgelig også fordi tiden bagefter har tegnet et billede hvor det studie ligesom viste vejen frem for hvordan er det vi skal lave... det er medicinafprøvning, men hele designet i den medicinafprøvning har vist vejen frem. Man har før undersøgt noget andet medicin indenfor det samme, hvor man har brugt et surrogatmål, og da der så var gået nogle år fandt man ud af at det kunne man overhovedet ikke bruge til noget. Og så kom det her studie som det første i den nye måde at undersøge det her på. Selvfølgelig er de begge to randomiserede, men det ene var med et surrogatmål, og det her er så med hardcore-endpoint, fraktur, hvor man siger at det er det man egentlig vil undersøge. Og så har den så dannet skole for hvordan man skal lave de her undersøgelser efterfølgende, og alle de præparater der er blevet undersøgt indenfor det medicinske område har fulgt den skabelon som den undersøgelse tegnede. Og den er veldesignet fra starten og det var en epidemiolog der satte sig ned og sagde, "Nu skal vi gøre sådan her", og så gjorde man sådan. Og man fik også det resultat man ønskede, altså den holdt vand. Så derfor har den dannet skole for hvordan man undersøger det. Det er den ene side af det, den anden side er så at det var noget medicin, altså man kan sige en ting er selvfølgelig at metoden er trendsættende, noget andet er så at det var en ny type medicin, hvor det var den første indenfor den gruppe medicin som var effektiv. Så der er to sider.

Question: *Hvad vejer tungest, synes du?*

Response: Det tror jeg ikke man kan sige, den ene er tungere end den anden.

Question: *Jeg tænker på i forhold til artikel (2) her at du har hhv. metodeudviklingen og så medicinen.*

Response: Jeg synes begge dele er vigtigt. Altså, metoden skal være i orden, det vejer selvfølgelig tungt at den er metodeskabende. Hvis man skal lægge noget vægt, så skal den være der. Men begge dele er meget vigtige synes jeg, også det at det er et nyt princip, en nyt behandlingsboldgade, at man kommer ind og rent faktisk finder noget der virker og som har vist sig i de næste 20 år faktisk at holde stik, i de undersøgelser der kommer efterfølgende.

Question: *Så metodeudvikling er i hvert fald noget der kan være høj kvalitet?*

Response: Ja.

Question: *Hvis vi så ser bort fra dem nu, og jeg så bare siger forskningskvalitet... hvad tænker du egentlig på når jeg siger det?*

Response: At folk opfører sig anstændigt og gør...

Question: *...det var første gang jeg hørte det... hvad mener du med anstændigt?*

Response: Anstændigt, der mener jeg også i forhold til metode. Du skal lave tingene med nogle ordentlige metoder, du skal ikke bruge en dårlig metode til at eftervise noget, du er nødt til at gøre det ordentligt. Det kan så selvfølgelig være dyrt og svært, men så må du stadigvæk være anstændig i det hvad du gør i din afrapportering, for næste trin det er jo så din afrapportering. Så skal du fortælle hvad du gør - og det er jo en selvfølgelighed. Men nu var jeg til møde i sidste uge om videnskabelig uredelighed, hvor der var et par eksempler, hvor folk ikke har fortalt hvad de gjorde. Og så lige pludselig tager fanden ved det. Så helt grundlæggende er det allervigtigst; man skal fortælle hvad man gør, så præcist og nøjagtig som mulig, og så skal man tænke sig om inden man begynder sådan man ikke kaster sig ud i noget, der ikke kan blive til noget.

Question: *Er der en stor fare for det, også i dag, eller er det kun i Danmark vi har sådan nogle fine protokoller?*

Response: Jamen den ene af de cases der blev trukket frem i sidste uge det var en artikel publiceret i Lancet, som var det rene snyd. Den var ikke fra Danmark, den var fra Norge, med en der havde siddet og opfundet data. Og den eneste grund til at de blev gennemskuet var fordi en af reviewerne vidste at den database der blev refereret til var nyere end det årstal der var blevet brugt i artiklen. Så der var en der havde løjet så åbenlyst at det var helt indlysende at det gik galt. Og sådan nogle sager de klistrer jo, og så er der jo Milenkowa, som jo er en træls sag. Og det der karakteriserer sådan nogle sager, det er jo nemlig manglende anstændighed, manglende ærlighed. Så sådan en åbenhed og ærlighed... og der synes jeg faktisk det BMJ gør med at reviewene er åbne, de er ikke skjulte. Navnene kommer ikke frem på reviewerne, men reviewernes kommentarer er tilgængelige. Og det synes jeg er med til at give en anstændighed til det. Den der åbenhed gør at der ikke er noget fordækt, ikke er noget skjult.

Question: *Hvor længe har det været det?*

Response: Det ved jeg ikke, det har jeg ikke styr på, men de er åbne.

Question: *Det er da relativt nyt, er det ikke?*

Response: Jo det er det, og jeg tror det er det eneste tidsskrift der gør det, jeg har ikke kendskab til andre.

Question: *Der findes nogle få, men det er ikke mange, det er det ikke. Det er det første*

tidsskrift af den størrelse jeg hører om.

Response: Ja, med den impact. Men det stimulerer til anstændighed, for du kan ikke tillade dig perfide kommentarer i dine reviews, at hagle den ned, for lige pludselig kan andre jo gå ind og se hvad der står. Så opfører folk sig mere anstændigt og ordentligt. Eller reelt. Nu siger jeg anstændigt, man skal være reel og åben. Jeg kan godt lide åbenhed også.

Question: *Hvad så hvis man skal ind og evaluere forskning, kan man det? Altså kvaliteten af forskning?*

Response: Ja så skal du jo beslutte dig for nogle kriterier for kvaliteten af forskning.

Question: *Har du nogle bud på sådan nogle kriterier?*

Response: Jamen det er vel ret enkelt, sådan grundlæggende. Du skal jo selvfølgelig definere din hypotese, så skal du have en metode der matcher din hypotese, og så skal du udføre det i henhold til det du har vedtaget, og så skal du afrapportere resultaterne som du finder dem, uden at prioritere i det. Man kan jo prioritere i sine resultater, så man kan veksle lidt med afrapporteringen. Men den er jo selvfølgelig svær, for tidsskrifterne vil jo kun rapportere det, der er nyt, eller de lægger vægt på det der er nyt, og det er en klassisk problemstilling. Der er heller ikke mange der gider at sidde og læse trivialiteter, eller negative fund.

Question: *Men hvad så egentlig med negative fund - er det dårlig forskning?*

Response: Nej det er det bestemt ikke. Hvis du får lov at afrapportere et negativt fund, så er det fordi det er gjort ordentligt. Der vil jeg sige at jeg vil forvente, at der i tidsskrifterne ligger en afvejning, hvor det er lettere at forkaste en dårlig undersøgelse med negative fund, så er det lettere at forkaste en dårlig undersøgelse. Hvis det er en god undersøgelse, så kan du bedre få det publiceret, for det første. For det andet er negative resultater mange ting, det kan jo være et interessant negativt fund, hvis alle nu har fundet en sammenhæng imellem to ting, og du så finder at der ikke er en sammenhæng, og dit design i virkeligheden er bedre end andres, altså din metode er bedre end andres, så er det jo et vigtigt negativt fund. Så er det jo ligefør det er positivt.

Question: *Det kan man så måske diskutere.*

Response: Du kan i hvert fald se der bliver publiceret mange negative fund omkring østrogener og hjertekarsygdomme, at "østrogen gav ikke øget forekomst af hjertekarsygdom", og så kom een ordentlig undersøgelse, et stort randomiseret forsøg, så blev det stoppet før tid, fordi der rent faktisk var en øget forekomst af blodpropper i hjertet hos kvinder i østrogenbehandling.

Question: *Du sagde også noget andet før, som... ja, det her som du kalder anstændighed, det var jo blandt andet noget med at man skulle skrive hvad man gjorde. Det er vel også et spørgsmål om reproducerbarhed?*

Response: Ja, og det er jo selvfølgelig at beskrive hvad du gør. Hvis du gør det i et laboratorium, så er det jo ret enkelt. "Så brugte jeg så mange molær af det og puttede i og efter så lang tid og ved den pH.", imens hvis det er mere bløde parametre, som f.eks. interview-undersøgelse, hvilke spørgsmål blev så stillet? Så bliver det mere diffust, og sværere at afrapportere. Og så må man jo så forholde sig så stringent til det som man kan. Og der bliver man nødt til at appellere til folks anstændighed, altså, "Jeg stoler på at du giver det referat, der bedst mulig repræsenterer det, der bliver snakket om til interviewene.", for du skal jo have det strømlinet på en eller anden måde.

Question: *Jeg tænker også på... jeg har hørt nogen nævne, at det ikke er alle tidsskrifter der giver særlig meget plads til metode. Er det noget du er stødt på?*

Response: Nej, det er det faktisk ikke.

Question: *Men er det ikke et problem i forhold til at kunne beskrive hvad man gør?*

Response: Jo, hvis man lægger begrænsninger på metodeafsnittet. Det har jeg faktisk ikke, jeg har lige fået en artikel tilbage, hvor de be'r om at få noget mere metode. Det synes jeg er rigtig godt.

Question: *Men du har ikke oplevet det andet?*

Response: Nej, det har jeg ikke.

Question: *Okay, godt.*

Response: Jeg synes så nogle tidsskrifter gør så det som jeg synes er fornuftigt, så nedskalerer de det, så det lige er en skrifttype minde, men det er der, og så må man tage forstørrelsesglasset frem og gå ind og læse det. Men jeg har ikke opdaget nogen der har bedt om at reducere metodeafsnittet.

Question: *Det kan være det er specifikt for et bestemt felt.*

Response: Vi publicerede faktisk en artikel her for et par år siden hvor der var 6-7 siders metode, der var i hvert fald rigtig meget metodeafsnit, og det var der ingen problemer i. Så nej, det har jeg ikke oplevet, og det tror jeg heller ikke anstændige tidsskrifter de gør.

Question: *Jeg ved ikke hvilke tidsskrifter det har været...*

Response: ...nej nej, men det er simpelthen helt basalt for ordentlig afrapportering af

dine forskningsresultater, dit metodeafsnit, det skal være ordentligt.

Question: *Er der sammenhæng imellem kvaliteten af forskningspublikationer, altså artikler, og så den forskning der ligger bag?*

Response: Om der er en association imellem de to ting? Det er der jo nok, det er der nok ingen tvivl om at der er, men du kan jo nok godt lave en elendig afrapportering af god forskning, selvfølgelig kan du det.

Question: *Men den anden vej rundt?*

Response: Den anden vej rundt? Det er jo straks sværere, hvis du har lavet noget crap forskning, og så gå ud og beskrive det som noget råd også...

Question: *...så er det stadig dårligt.*

Response: ...så er det stadig dårligt. Forskning bliver ikke god. Du kan selvfølgelig vælge at dele det ind og sige, en elendig forskning, dårligt afrapporteret, er selvfølgelig ikke dårligere end elendig forskning korrekt afrapporteret.

Question: *Men det er stadig ikke god forskning?*

Response: Det er stadig ikke god forskning. Men god forskning dårlig afrapporteret, det ved jeg ikke...

Question: *...det er synd*

Response: Ja! Det er synd og skam, det er simpelthen for åndssvagt. Så dum kan man ikke tillade sig at være. Det er frygtelig synd.

Question: *Ja, et eller andet sted. Hvad med sådan noget som studiedesign. Sådan noget som klinisk randomiserede forsøg, kasuistikker og så videre...*

Response: Ja, den der prioriterede liste.

Question: *Ja, lige netop. Er der en sammenhæng imellem det og så kvalitet?*

Response: Jamen det er der selvfølgelig. Dårligt design det er dårlig kvalitet. Et dårligt designet studie risikerer jo at vildlede i stedet for at vejlede. Så derfor må det per definition være dårlig kvalitet.

Question: *Hvad afgør som om det er et dårligt design?*

Response: Jamen jeg synes den prioriterede liste der er angivet i epidemiologisk bøger eller i forskningsmetode, det synes jeg... det er klart, case reports er godt til at få idéer,

men du kan ikke bruge det til noget som helst. Epidemiologi er også meget godt til at få idéer, men du skal lave randomiserede studier før du kan stole på resultaterne. Det er for eksempel den vi har med østrogen og hjertekarsygdom. Der var mange case-control-studier, og observationelle studier og retrospektive studier der viste en positiv effekt, og da man endelig lavede det randomiserede studie så var det præcis det modsatte der var tilfældet. Man troede der var den her beskyttende effekt, og så var der en 20 – 25% øget risiko da man endelig lavede det randomiserede studie. Så det må man sige, den der rangorden af kvalitetsdesign, den er vigtig. Det skal man have i bagehovedet, når man læser artikler. Det er nødvendigt.

Question: *Er det så som kliniker du vurderer det? Eller er det...*

Response: Det er begge dele. Altså kliniker og forskeren, det bør man ikke adskille. En god kliniker er i stand til at forholde sig til den litteratur han bliver præsenteret for, og ligesådan forskeren skal også kunne se den kliniske gevinst af det, for at sige, hvad er det værd at bruge krudt på? Synes jeg. Og det leder frem til en helt anden diskussion, som jeg synes er vigtig, nemlig at lægers, eller forskeres, stillinger, hvis man er ren forsker, så kan det måske være meget godt i nogle sammenhænge, men man risikerer at miste jordforbindelsen. Man får flere relevante idéer, hvis man har en blandet ansættelse, hvor man både har forskning og klinik.

Question: *Er det specifikt for dit felt, eller er det helt generelt?*

Response: Det tror jeg gælder alle felter. En infektionsmediciner, f.eks., hvis det er en der kun sidder og kigger i petriskåle, og analyserer bakterier, i stedet for en der har lidt klinik også og ligesom holder fast i hvad der er relevant. Man kan selvfølgelig argumentere for, at der skal være begge dele, det kan man selvfølgelig godt. Jeg synes det er en god idé at have begge berøringsflader, for at holde sammen på tingene.

Question: *Så som kliniker bør man også være forsker?*

Response: Det synes jeg faktisk, så bliver du mere... ikke kritisk, det er måske et forkert ord, for det er jo ikke for at kritisere den litteratur man bliver præsenteret for. Men vi går jo og behandler patienter, og for at være sikker på at det vi går og behandler på ikke er fisket ud af den blå luft, jamen så skal vi jo kunne læse de undersøgelser der er, med en viden om hvad troværdigheden er af de enkelte undersøgelser. Fordi vi ved jo godt det medicinalfirmaerne kommer og præsenterer for os det er selvfølgelig vægtet ud fra hvad de synes er relevant for deres præparat, og så kan de godt glemme at præsentere nogle andre undersøgelser. Der skal vi være i stand til selv at kunne gå ind og finde dem. Det kan jeg godt komme med eksempler på, men det er ikke det det handler om.

Question: *Nej, ikke specifikt, men jeg kan godt følge dig uanset. Ved du tilfældigvis hvor ofte de her to artikler er citeret?*

Response: Nej, det gør jeg ikke. Men jeg ved at artikel 1 var indsendt fem gange inden et tidsskrift antog den, fordi alle mente, at det kunne ikke passe. Og det passede så, og da den så var blevet antaget, var der adskillige andre der eftergjorde det, og fandt det samme, og derfor er den stærkt citeret. Og altså, Peter er jo vores her i huset, han er jo blevet verdensberømt på, han laver mange gode ting, men det her er en af de ting der har været med til at slå hans navn fast som idémaker indenfor den her verden. Han kan komme op med nye idéer som andre folk ikke kommer på. Folk de tænker, "det passer nok ikke det her", nej, det ved jeg ikke hvad de tænker, men Peter kommer op med en idé, forfølger den, siger "Det er jo sådan det er, det må jeg skrive om", det er jo sådan det er. Så derfor er den stærkt citeret. Den anden er selvfølgelig også stærkt citeret, jeg ved ikke hvor mange gange, men 100 citater, slag på tasken, hvis det kan gøre det. Nej, det kan ikke gøre det, for den er helt grundlæggende i al osteoporosebehandlingsforskning der er kommet de sidste 15 år.

Question: *Det er også mest for at jeg ved om du har været inde og se på det inden du valgte dem, men det kunne jeg jo så næsten også regne ud du ikke havde, eftersom vi sad her og fandt dem.*

Response: Nej, det er ikke derfor, de er fisket ud af baghovedet.

Question: *Godt... det sidste jeg har er nogle begreber jeg vil nævne, og så vil jeg gerne at du bare svarer det første du tænker på når du hører det, selvfølgelig i konteksten forskningskvalitet. Og det første jeg har, det er ny viden.*

Response: Forskningskvalitet og ny viden, jamen de to ting de hænger jo stærkt sammen synes jeg. Ny viden... du kan godt have ny viden fra ting der har dårlig forskningskvalitet, men så er du nødt til at eftergøre det. Men hvis du kombinerer de to ting; en god, velgennemført undersøgelse, altså en kvalitetsundersøgelse, der præsenterer ny viden, den får stor impact, den slår hårdt igennem. Så du kan godt lave noget crap-forskning, der bidrager med ny viden, men omvendt... implementeringen af den ny viden forudsætter... hvis den er dårlig kvalitet er den hypotesegenererende, hvis det er god kvalitet så er det troværdigt. Det skal være god kvalitet for at det kan være troværdigt. Men ny viden vil altid på en eller anden måde give en opmærksomhed. Men der skal man jo så passe på... den case der var der fra Norge, det var NSAID-præparater og oral cancer, det var jo en ny idé, det var ny viden, og det var det rene fup.

Question: *Ja, de falder lidt udenfor kategori.*

Response: De falder lidt udenfor kategorien, ja, men det er jo sjovt nok at tænke på, hvis der kommer noget som er helt ny viden, som ikke passer med noget som helst, så er du jo nødt til at tænke dig om en ekstra gang og sige, "kan det her nu virkelig passe? Er det nu virkelig sandt?". Så man kan sige ny viden forudsætter måske i virkeligheden god kvalitet. Det kunne man godt argumentere. Både ny viden er kvalitet i sig selv, det

er en kvalitet at producere ny viden, men for at den kan være troværdig, så skal det være god kvalitetsforskning der ligger bag ved.

Question: *Hvad med peer review?*

Response: Den proces er helt afgørende for troværdigheden. Troværdighed, det vil være mit første ord. Det må godt være en åben peer review proces, men dem der laver peer reviewene de må godt være unavngivne. Men selve reviewene... jeg kan godt lide princippet omkring det åbne. Men jeg mener peer review er helt centralt for forskningskvalitet og troværdighed. Og det er også derfor der er nogle af de der online tidsskrifter der laver noget der hedder reader-peer-review, du skriver din artikel, så publicerer de den, og så må læserne bagefter kritisere den.

Question: *Du ligner ikke en der er fan.*

Response: Arh, det kan jeg ikke gå ind for. Det er da fuldstændig... du kan jo skrive hvad som helst, der er jo ingen der går ind og kontrollerer om det er rigtigt. Der er ikke nogen der går ind og ser om det har gang på jord, om der er sammenhæng i det du skriver, om det giver mening, det synes jeg godt nok... så kunne man ligeså godt læse Ekstra Bladet.

Question: *Okay... men læserne er vel også forskerne?*

Response: Det kan du vælge, men du kan også vælge selv at oprette et alias og så selv at gå ind og lave en reader-peer-review proces på det du selv har publiceret. Den risiko er der jo i den proces, så det mener jeg er et helt centralt begreb.

Question: *Hvad med tværfaglighed?*

Response: Spændende. Tværfaglighed kan give nogle nye idéer. Det gør det ikke altid, men det kan give basis for nogle nye idéer. Og det vil sige så kan man komme frem til nogle spændende pointer, som man måske ikke var kommet på hvis ikke man var gået på tværs af grænserne.

Question: *Hvad tænker du på, når jeg siger tværfaglighed?*

Response: Jamen på tværs af specialer. Det er det, jeg umiddelbart tænker.

Question: *Okay. Hvad med layout af artikler?*

Response: Ja. Glæde. Jamen ved du hvad, en artikel i et godt layout, jeg bliver så glad når jeg ser det. Det er sjovt, nogle tidsskrifter er bedre til det end andre, og skal jeg være helt ærlig, så kan jeg faktisk godt finde på at vælge et tidsskrift, der har et godt layout. Altså et indbydende layout, hvor man kan se tingene hænger sammen, figurerne er trykt i en... ikke puttet ned i et hjørne i mikroskopisk størrelse, men sådan op hvor man rent

faktisk kan se hvad der står under dem, og det passer med teksten, det står ved siden af den del af teksten hvor det hører til. Det er simpelthen... glæde. Og skal jeg trække den lidt, så kan jeg godt finde på, hvis der er to tidsskrifter med ens impact og ens relevans for det jeg har, så vælger jeg den med det bedste layout.

Question: *Nu siger du jo faktisk selv det næste, nemlig impact.*

Response: Jamen impact factor bliver jo brugt af bevillingsgiverne til at vurdere hvem der skal have bevillinger. Hvis du har lavet et antal artikler i et fantastisk godt tidsskrift, så har du bedre chancer for at få bevillinger. Og det vil sige at de går ind og kigger direkte på impact factor, og det smitter selvfølgelig af på valget også. Så derfor bliver man nødt til, når man vælger tidsskrift at publicere i, så at vælge det.

Question: *Men er der sammenhæng imellem det, og kvalitet?*

Response: Ikke altid. Den der gyselige artikel, der blev kritiseret sønder og sammen, den kom i Lancet. Men den var ikke gået i BMJ med de åbne reviews, det tror jeg faktisk ikke. Man kunne få den tanke, at det var fordi nogle af de der meget high-impact tidsskrifter angler efter noget, der kan komme i pressen, noget der kan give noget PR. Altså, de har en sekundær dagsorden. En ikke-videnskabelig dagsorden. Den mistanke kunne man godt have. Og det mener jeg faktisk der er nogle af dem der skriver inde på deres hjemmesider også. Men impact factor betyder selvfølgelig noget. Men det var ikke derhen du spurgte ad...

Question: *Nej, det var mere hvis man nu ser bort fra de her særtilfælde, er der en generel sammenhæng imellem impact factors og kvaliteten af de artikler der er i de tidsskrifter?*

Response: Måske, fordi low-impact tidsskrifter de kan også godt have... jeg sidder sådan og tænker på artikler jeg har læst i low-impact tidsskrifter, så synes jeg det er udmærket kvalitet, dem jeg lige sidder og tænker på. Så tidsskrifter med lavere impact factor publicerer også i god kvalitet. Selvfølgelig sådan nogle resultater som er mindre spektakulære og mindre banebrydende. Men derfor er forskningen alligevel af god kvalitet. Men det er klart, du kan vende den om og sige; undersøgelser af dårlig kvalitet, har en bedre chance i tidsskrifter med lavere impact. Hvis du har lavet en dårligt designet undersøgelse, så har du en bedre chance for at få den publiceret i et lavere impact tidsskrift.

Question: *Du er sådan set inde på noget jeg synes er meget interessant her. Du siger, det kan godt være god kvalitet, uden at det nødvendigvis er banebrydende. Er det måske i virkeligheden det impact factor er - den siger noget om hvor banebrydende forskningen er?*

Response: Det tror jeg, den hælder mere i den retning.

Question: *Og kvalitet forstår jeg mere, at det opfatter du som en slags...*

Response: I den her sammenhæng af den underliggende forskning.

Question: *Ja, og det er mere noget form, at det er reproducerbart, at man har gjort det ordentligt, har valgt et design, der passer til problemet...*

Response: Præcis, det er kvalitet i min...

Question: *Mmmm. Det sidste begreb jeg har er citationer.*

Response: Ja, men det er faktisk ikke noget jeg rigtig kigger på.

Question: *Men er der en sammenhæng imellem dem og kvalitet?*

Response: Jeg tror egentlig mere det hænger på om det er nyskabende. Nyskabende ting bliver jo citeret mere end kvalitetsting... men det er jo så, jaeh, og så sidder man tilbage og tænker... hvis der nu er lavet tre artikler om det samme emne, og de finder det samme, så er jeg faktisk ikke sikker på det er den med kvalitet der nødvendigvis bliver citeret mest. Det burde det være, men hvorfor egentlig det? Personligt ville jeg selvfølgelig vælge at citere den der var bedst, metodemæssigt og afrapportermæssigt. Nej, jeg har ikke evidens for at sige andet.

Question: *Men du har en mistanke?*

Response: Man kunne godt få tanken, at der... men hvad vælger folk så efter? Det kan være de vælger efter hvilken en der er blevet citeret før, og så tænker de at den er nok god at citere, ergo kan jeg også bruge den, så kan jeg nøjes med at læse referencelisten på andre artikler, eller de lignende artikler, som bruger den. Så er jeg fri for selv at bestille artiklen hjem og læse den. Det er jo, det er dårlig kvalitet.

Question: *Men ikke uset?*

Response: Men ikke sjældent. Man kunne godt forestille sig, at det er sket flere gange. Og det kan også være svært... det kræver en indsats at gå ind og finde alle artiklerne og læse dem alle igennem. Men sådan lige at skimme metodeafsnittet, og se om det nu passer til det man har undersøgt, er det nogle ordentlige laboratoriemetoder, er det et ordentligt design, er det et ordentligt antal personer der er med, der i virkeligheden giver den styrke der skal til... det er nok en forudsætning for at det bliver lavet ordentligt. Også en forudsætning for at det kan blive citeret.

Question: *Jamen det var sådan set det jeg havde så. Medmindre du har noget du mangler at sige.*

Response: Åh, hvad kan man sige? Nej, jeg tror vi er kommet rundt om det der var hen ad vejen.

A.2.3. Interview reference: M04

Date: Nov 08, 2011

Time: 14:00

Question: *Du har taget to artikler med, og jeg har bedt dig om at tage dem med på baggrund af at du mente det var særlig høj forskningskvalitet? Har du lyst til at fortælle mig noget om hvad du mener det er ved de her artikler der er særlig høj kvalitet?*

Response: Jamen det kan være jeg har misforstået dig lidt så, jeg tænkte det skulle være nogle særligt gode artikler.

Question: *Jamen det vil jeg også gerne høre noget om.*

Response: Det er selvfølgelig også et bidrag til forskning. Men vi kan tage den her først (1), af Rosemary Fischer... jeg skal måske også sige, at jeg har valgt nogle artikler som er relevante for mit meget snævre forskningsprojekt, så det er jo ikke sådan at jeg har været ude og kigge på alt hvad der er skrevet i denne verden. Men for mit projekt så synes jeg at denne her artikel 1 er af høj kvalitet, fordi det er nogle meget kompetente folk der har lavet det, så der har jeg jo det handicap, eller det bias, at jeg kender førsteforfatteren. Men også fordi, hvis jeg ikke havde gjort det, og hvis jeg kendte til mit forskningsprojekt, så kunne jeg se, at det her er et område, der er rigtig svært at forske i, og det er rigtig svært at få klinisk materiale, og hun har noget godt klinisk materiale hun har forsket i. Og så har hun forsket med gode metoder, og forskellige metoder, og så gengiver hun sine data klart og uden at snyde, og det er måske i virkeligheden det der er rigtig vigtigt for mig; at man ikke har indtryk af, og det tror jeg heller ikke er tilfældet, at hun har skrevet det her for at få en yderligere publikation. Hun har simpelthen skrevet fordi her er et budskab der er relevant og vigtigt, og som vi ikke kan fange på andre måder. Og det gør det til god forskning.

Question: *Nu siger du hun bruger flere metoder, hvor almindeligt er det?*

Response: Hvor almindeligt er det i det, der bliver publiceret?

Question: *Ja.*

Response: Nogle gange kan man jo se at man ikke har gjort det, der får man indtryk af, at folk gerne vil have mange publikationer.

Question: *Okay, og så tager de måske en publikation per metode?*

Response: Lige præcis. Og der har hun mere sat sig for at besvare spørgsmålet. hun gør en observation som er usædvanlig, og som kan være med til at lære os noget om hvordan

verden er skruet sammen, og den observation angriber hun så med mange forskellige metoder, for at udhente mest mulig information.

Question: *Når jeg sådan lige ser på titlen, er det korrekt forstået at det er en enkelt case hun ser på?*

Response: Ja, det er en enkelt case, men det er de her skæve graviditeter, som er så enormt sjældne, at her giver det ikke mening at tale om store datasæt, dobbelt-blindet og overkrydset, her er det at gå ud og se hvad naturen har lavet af eksperimenter. Og se hvad vi kan lære af det. Og der er hun så faldet over en, som ingen andre har set, nogensinde, i verden. Jeg har set noget der ligner nu, og som vi kan skrive noget spændende om, men ikke er magen til. Men det er relevant at hive een case frem og angribe den med mange metoder. Så det gør det ikke til dårlig videnskab at hun starter med en case.

Question: *Ja. Hvad med den anden?*

Response: Jamen det er jo så i virkeligheden ikke forskning det her, kan man sige, for det er et review, men det er et review skrevet af en meget kompetent person, og hun referer også til en del af sine egne data, så hvis du tager den her oversigtsartikel og ser hvad der er i den, så er der masser af forskning i den, inklusive noget af hendes eget. Men jeg synes den er rigtig god, fordi den ikke er... dels fordi den er skrevet dækkende, altså mange reviews bliver skrevet, altså jeg er jo i genetik, så finder man reviews om et eller andet emne, så kan man se, jamen den er skrevet for børnelæger, eller den er skrevet for kirurger... og så er den rigtig let tilgængelig og næsten lærebogsagtig, men ikke særlig informativ hvis man ved noget om emnet. Men den her er skrevet af en genetiker, for genetikere, så hun er grundig. Og igen er den ikke... man har ikke fornemmelsen af at hun vil sælge noget, så det er faktisk rigtig lødig information.

Question: *Hvad mener du med det? Hvad mener du med, at hun ikke vil sælge noget?*

Response: Altså nogle gange, sælge sit eget image vel. Hun er ikke i gang med at lave noget for at svine noget papir til, hun er igang med at give noget information til nogle mennesker som kan bruge det til noget.

Question: *Når du siger, det ikke er forskning...*

Response: Det er fordi hun præsenterer ikke upublicerede data her, alt det hun skriver om her det er tidligere publiceret, så på den måde er det jo ikke en original publikation.

Question: *Nej, den er ikke original. Men er det ikke stadigvæk forskning?*

Response: Joeh, den handler om forskning.

Question: *Men der kommer ikke noget nyt frem i den?*

Response: Der er ikke nogen, altså, alt det her ville man kunne have fundet selv, hvis man har tyret litteraturen så godt igennem, som hun har. Hun samler det jo op for mig, så jeg ikke behøver tyre litteraturen igennem. Det kalder man så ikke forskning i min verden. Det kalder man nok forskning i andres verden, det skal jeg ikke kunne sige, men i min verden skal man generere noget, som ingen andre kunne, med mindre de også har været ude og lave eksperimenterne.

Question: *Okay, ja. Så tænkte jeg på, et rent formspørgsmål, er det et struktureret review, eller ustruktureret?*

Response: Ja, det er meget struktureret. Og man kan simpelthen, hvis jeg gerne vil vide noget om genetik, så kan jeg bladre hen til den overskrift, hvis jeg vil vide noget om klinik, kan jeg bladre derhen. Så det er meget struktureret, igen i modsætning til hvis hun skulle skrive en lille historie for børnelæger, så ville hun have skrevet det essay-agtigt, ikke?

Question: *Okay, godt. Hvis vi så skal se på forskningskvalitet helt generelt, hvis vi ser bort fra de her to artikler, eventuelt også ser bort fra artikler, det må du selv om, hvad tænker du så på, når jeg siger forskningskvalitet?*

Response: Jamen det kan man vel ikke sige sådan ud i den blå luft. I hvilken sammenhæng? God forskningskvalitet, mener du?

Question: *Ja, især...*

Response: Jamen det kunne være, at man bibringer verden noget viden, der ikke var der før. Men man kunne jo også argumentere for, at man bibragte verden noget relevant viden. Noget nyttig viden, man kunne bibringe noget viden på en økonomisk måde... man kan jo sjuske med data, og ikke hive al den viden ud af det, der er mulig, eller indsamle alt for mange data, og kunne godt have konkluderet på et tidligere tidspunkt. Så man kan også tale om at være økonomisk, så kvalitet kan jo være rigtig mange ting.

Question: *Der er en af de ting du siger, som jeg synes er spændende i hvert fald. Når du siger relevant viden - findes der irrelevant viden?*

Response: Ja, det gør der i mit fag. Jeg er jo læge, så hvis jeg skal se på at jeg også er del af et system, hvor man gerne skulle holde raske mennesker så raske som mulige og gøre syge mennesker raske, så er der jo helt klart noget som er mere relevant end andet. Så der er jeg nok lidt handikappet af min baggrund. Når vi forsker i mit fag, så tænker vi jo tit, "Hvad kan det bruges til?", også selvom vi er helt nede og kradse i grundforskning, så svæver den stadig rundt; Kunne man forestille sig, at det her på et eller andet tidspunkt kunne komme nogle mennesker til gavn?

Question: *Men det kan ikke simpelthen bare have så lange fremtidsudsigter, at det er fordi*

vi ikke kan se det?

Response: Jo jo jojo, og jeg mener også at grundforskning i sig selv er noget der skal være plads til i vores samfund, også der hvor man overhovedet ikke kan se nogen nytte af det her og nu. Men derfor kan man jo godt vælge at have det synspunkt, at det er en kvalitet, at det er noget der umiddelbart er en nytte i.

Question: *Så kunne man måske sige, at som en slags klinisk forskningskvalitet vil være, at det har en umiddelbar relevans?*

Response: Ja, det kunne man sige, ja.

Question: *Er der så også en sammenhæng imellem kvaliteten af den forskning man udfører, og så kvaliteten af de artikler der kommer ud af det?*

Response: Det vil jeg da tro. Det vil jeg da håbe, det er ikke noget jeg har kigget på, men det vil man da egentlig antage. Men jo ikke nødvendigvis, man kan jo godt være en sjusk med sine data, at man faktisk har gode råvarer, men man får det ikke præsenteret ordentligt.

Question: *Hvad med den anden vej rundt?*

Response: Jamen de kan vel aldrig blive bedre end de var. Så det ville for mig være lidt volapyk.

Question: *Så hvis man gik ind og vurderede en bestemt artikel og vurderede at det var virkelig høj forskningskvalitet, så kan man også formode, at studiet der ligger bag, er også høj forskningskvalitet? Medmindre vi snakker om fusk og snyd, men hvis vi nu ser bort fra det...*

Response: Nej, det vil jeg sige det kan du jo netop ikke, nå ja, jo, okay, hvis den er letlæselig... ja. Nej, jeg trækker det hele i mig igen, det må være sådan. Hvis artiklen fremtræder fornuftig, så er det svært at forestille sig, at der ikke ligger fornuftige data bag, medmindre man har sjusket, ja. Det må du have ret i.

Question: *Hvis man skal ind og evaluere forskningen, kvaliteten af forskningen, kender du til nogle metoder til at gøre det? Kan man overhovedet det i din optik?*

Response: Jamen man kan det jo med talforskning, man kan jo måle signifikansniveauer og sandsynligheden for fejlagtige konklusioner til den ene og den anden side, hvis det handler om talforskning.

Question: *Ja, jeg tænker egentlig mere på hvis man skulle ind og vurdere om den ene artikel var bedre forskning end den anden artikel. Det er jo vældig populært at gøre.*

Response: Jamen det kunne man jo godt, f.eks. i den her artikel her (1), der konkluderer hun at der er et kromosom 11 for meget, som stammer fra mor, men reelt kan man godt se at hun har haft svært ved at hive sine data i land, for når man kigger ned ad kromosom 11, så er det ikke alle steder hvor man rent faktisk kan se hvad der er arvet fra far og mor, så på et eller andet tidspunkt så tænker man , nu har jeg nok data, nu konkluderer jeg. Og der kan man jo godt gå ind og kigge kritisk på, hvor solide data havde forfatterne før de gik ind og konkluderede. Eller, giv et eksempel på hvor mange metoder der er brugt, ved at kigge på den med mange metoder. Og der kan man jo godt lave en eller anden form for skala, hvis man nu ville. Og det kommer jo helt an på emnet, om der er nogle metoder som man bør bruges, nogle golden standards.

Question: *Så det jeg hører dig sige er, at man simpelthen er nødt til at gå ind og vurdere den enkelte artikel i forhold til emne og metodevalg?*

Response: Helt klart. Men man kunne selvfølgelig godt, hvis man ville lave nogle generelle retningslinier finde nogle ting, man kunne godt beskrive det i generelle termer. Man skal sætte sig ind i emnet og hvilke metoder der findes, og findes der så en eller flere? Findes der 100? Skal man så bruge alle eller de tre bedste? En anden ting er jo også om man står på skuldrene af de rigtige? Har man været den forrige litteratur igennem og er det fornuftigt det man laver i forhold til det andre har lavet, og sætter man sine resultater fornuftigt ind i en sammenhæng i forhold til hvad man i øvrigt ved fra andre data? Sådan nogle ting kunne man selvfølgelig beskrive i generelle termer, men skal man jo selvfølgelig, når man skal bruge det på et konkret forskningsprojekt, gå ind i det og se hvad findes der egentlig af metoder og anden viden, osv.

Question: *Så det du siger, bl.a., er faktisk at man ud fra referencelisten kan få et fingerpeg om, om det er et ordentligt arbejde eller ej...*

Response: ...hvis man kender til emnet, ja.

Question: *...hvis de refererer til alle de relevante, er der en større sandsynlighed for at de også har gjort det fornuftigt, end hvis det bare er tilfældige referencer?*

Response: Ja. Og det gør man jo også selv hvis man er reviewer. Som reviewer vil man jo straks studse over, hvis de har refereret til nogen besynderlige ting, og det korrelerer også tit med hvordan de i øvrigt skriver.

Question: *Hvad med sådan noget som evidensniveau? Du sagde f.eks. at den her ovre (1) var en kasuistik, men den er stadigvæk god kvalitet. Jeg har hørt mange drage en slags parallel, ikke en direkte parallel, men en slags parallel imellem kvalitet og evidens. Der ville den her jo så falde lidt udenfor, men er du sådan i øvrigt enig?*

Response: Hvorfor mener du at den falder udenfor?

Question: *Jo, fordi hvis man skulle købe den med at højeste kvalitet var meta-analyser og så dobbeltblindede, randomiserede forsøg, så ender kasuistikkerne jo ret langt nede i bunden.*

Response: Men det kommer jo an på hvilket fænomen man gerne vil beskrive. Hvis du gerne vil beskrive Big Bang, f.eks., så giver det jo ikke ret meget mening at tænke i dobbeltblindet, osv., den er jo fænomenologisk forskning, nemlig hvad forekommer der i verden. Og det er det, det er, det her. Altså, kan man leve, er det foreneligt med liv at ha' den her kromosombesætning, det kan man ikke... så skal man i hvert fald forestille sig noget meget specielt, enten uetisk, eller flytte det over på en dyremodel, hvis man skal til at gøre det her eksperimentelt. Og det giver vel ikke så frygtelig meget mening at snakke om dobbelt-blind, altså, så taler du om interventionsforskning, altså, vi har virkeligheden, så går vi ud og piller ved den, og så ser vi hvad der sker. Og så piller vi ved den på den ene måde, på den anden måde, og så ser vi hvad der sker. Det her er observationel forskning. Vi ser hvordan verden er skruet sammen.

Question: *Er det så fordi der er flere forskellige måder at se på evidens?*

Response: Ja, det vil sige, hun kommer jo med nogle konklusioner, og der kunne jeg jo gå i rette med hende hvis jeg havde lyst, sige har du nu ordentlig bevisførelse for det du konkluderer... det vil jeg jo så sige er evidens. Hvad er egentlig din evidens for at man godt kan have sådan en mola-graviditet, på trods af man har et matalt kromosom 11? Og så kan hun sige at hun kan se med den metode at det var matalt, og hun kan se med den metode, og med den metode, og jeg kan sige, "ja ja, det er meget godt med dig, men ned ad på kromosom 11 der var din markørtæthed ikke så stor.", der kan man jo godt diskutere evidens uden man har talt om hvor mange af den her slags biologiske fænomener hun har kigget på. Hun har kun kigget på eet biologisk fænomen, men med mange metoder. Det vil jo i mit hovede også være en slags evidens. Så kommer hun og siger, "jeg synes det lignede et mor-kromosom og ikke et far-kromosom", så kan man sige, det er en meget tynd evidens hun kommer med. Men kommer hun med den metode, den metode, den metode, da var det et mor-kromosom, så er det en tung evidens. Så jeg tænker jo mere på ordet evidens i ordets egentlige forstand.

Question: *Ja, og det kan jeg sagtens høre. Det er bare ikke i den forstand de fleste bruger det, når man nævner evidens overfor en almindelig kliniker på sygehuset, jamen så vil de jo som regel forstå det her med det randomiserede forsøg, og så derned af i niveauet, og der ligger kasuistikkerne ikke helt nede i bunden, men ret langt.*

Response: Men så er det måske forkert at kalde det en kasuistik, så skal man måske kalde det observationel forskning?

Question: *Ja, det kan godt være, at det måske er mere korrekt.*

Response: Eller også er det fordi mit sigte jo ikke er umiddelbart klinisk.

Question: *Ja...*

Response: ...det her skal ikke bruges til at behandle næste patient anderledes. Det her er grundvidenskab, hvordan er verden skruet sammen.

Question: *Ja, så derfor er det måske noget lidt andet man skal se på der.*

Response: Du ville jo heller ikke lave den slags evidens hvis du sad og talte med en fysiker.

Question: *Nej, netop. Og derfor kan jeg også følge dig i, at du sagtens kan vurdere en kasuistik som høj kvalitet, hvorimod en kliniker der måske skal til at overveje en ny behandling, måske ikke kan bruge en kasuistik.*

Response: Han ville jo måske bruge en kasuistik som et afsæt til at lave sit kotrollerede, kliniske forsøg.

Question: *Netop. Så har jeg egentlig kun nogle ganske få ting tilbage. Jeg har en slags brainstorm hvor jeg vil nævne nogle begreber, som jeg gerne vil høre din umiddelbare reaktion på.*

Response: Som den jeg er, eller skal jeg nu lade som om jeg er en meget klinisk kliniker?

Question: *Nej, du skal endelig sige det som du er. Fordi, klinikere er på en måde interessante, men det er grundforskningen jo også, og jeg vil sådan set gerne høre det hele. Og jeg har snakket med masser af klinikere, så.. Men det første begreb jeg har, det... du må selvfølgelig meget gerne sætte det i konteksten forskningskvalitet, men det første begreb det er ny viden.*

Response: Det har da en rigtig høj relevans. Hvis forskning ikke genererer ny viden, så er det jo indholdsforladt, tomt, meningsforladt.

Question: *Og det findes?*

Response: Ja, man kan da sagtens gentage nogle andres... altså hvis man nu bare vil have sit navn på nogle publikationer. Så gøre et eller andet rutinepræget og se om man kan få nogen til at trykke det.

Question: *Du lyder som om det er noget du støder på af og til?*

Response: Nej, men hvis man skal finde et forskningsprojekt for en student, f.eks., så er det jo fristende. Hvis man har et projekt hvor man mere eller mindre kan forudse hvad der kommer ud af det.

Question: *Tværfaglighed?*

Response: Det kommer vel an på emnet. Det er ikke et mål i sig selv for mig.

Question: *Det kan være nyttigt hvis det er nyttigt, men ellers så?*

Response: Jeg må sige, det må være fuldstændig underordnet, det må være at man laver noget ordentlig forskning, at de data der kommer ud af det er ny viden. Men hvordan man kom derhen synes jeg ikke er relevant. Men det vil ofte være tværfagligt.

Question: *Citationer?*

Response: Hvad mener du?

Question: *Feks. hvor mange gange en artikel er blevet citeret.*

Response: I forhold til forskningskvalitet?

Question: *Ja. Er der en sammenhæng imellem det?*

Response: Det er der sandsynligvis, hvis du tager alting. Men der er jo enorm bias i forhold til hvad emnet er. Jeg forsker i et ekstremt lille emne. Og jeg har lige været til verdenskongres, vi var 132, og deraf var de fleste klinikere. Vi er fire genetikere, og vi er reviewere for hinandens artikler, vi bliver aldrig citeret ret meget for den slags. Til gengæld har jeg været een ud af hundrede forfattere på en artikel om brystcancer gener, og den bliver jo citeret. Men det er jo lidt snyd, så citationer er et mål, men det er godt nok meget kontekstafhængigt, kan man sige.

Question: *Men hvis man nu holder sig indenfor et bestemt område, kan man så godt bruge dem til noget?*

Response: Ja, det må man sige. Men så skal området være specifikt.

Question: *Det næste begreb det er impact.*

Response: Altså impact factor for tidsskrifter, eller hvad tænker du på? Det er det ord der kommer ind i mit hovede.

Question: *Og det er helt i orden. Man kan fortolke det på flere måder, og hvis det er sådan du fortolker det, er det dét jeg gerne vil høre.*

Response: Altså jeg vil sige at den her artikel (1) burde få enorm impact, for den er formentlig med til at pege på hvor et gen, som har afgørende betydning for at mennesker overhovedet bliver mennesker, er lokaliseret. Den har ikke fået det endnu, men den får det. Den har impact. Men jeg tænker mest på ordet som impact factor for tidsskrifter, og som jo er vigtigt når man skal promovere sig selv, have den næste stilling, have nogle

forskningsbevillinger, så derfor er den jo en løftestang. Det er en kommunikationsform i den verden jeg også begår mig i, når jeg skal have nogle penge til min forskning.

Question: *Men er der en sammenhæng imellem impact factors på et tidsskrift, og så kvaliteten af den forskning der bliver publiceret deri?*

Response: Det er der sandsynligvis, i det store træk. Men der er også enorme andre kilder til hvad der bliver impact factor. F.eks. er der jo visse kliniske tidsskrifter, som bliver meget citeret. Men altså, det er jo en lang diskussion, hvad der er den bedste målemetode. Og man bliver nødt til at bøje sig for at det jo stadig er det sprog der tales. Så det gør indtryk på folk hvis der er tale om høj impact factor.

Question: *Layout af artikler?*

Response: What?

Question: *Der er ikke lige noget der?*

Response: Nej, det vil jeg sige er totalt irrelevant.

Question: *Helt i orden. Og det sidste begreb det er peer review.*

Response: Jamen det er jo et plus-ord, det er et buzz-ord. Jamen, det er jo godt. Nu er jeg ikke så fortrolig med ikke-peer-review, for mig er det jo sådan en selvfølgelighed. Men jeg kan måske sige noget komplementært, nemlig at vi i nogle, også i universitetssammenhænge bliver opfordret til også at publicere populærvidenskabeligt, og hvor vi også bliver målt på det. Men jeg synes da peer review er rigtig vigtigt, fordi, det kender jeg igen fra at være reviewer... der er godt nok mange interessante forslag til hvad man burde skrive, også noget som er åbenlyst vās, selv for en ret ukyndig. Så jeg synes peer review er ret vigtigt. Så det er en kvalitetskontrollant.

Question: *Er det minimum, eller er det... altså kan man sige at peer review er minimum for at man kan kalde det forskning? Eller sætter det en acceptabel standard, eller bare lige så det går an?*

Response: Jeg har svært ved at se at man... det synes jeg er svært at sige. Jeg synes jo peer review er rigtig godt, men jeg kan godt se at man jo godt kan lægge ovenpå, stille højere krav end peer review. Men igen, så bliver ting jo nogle gange lidt firkantede, så afhængig af hvilken forskning man laver, så er det jo lidt forskelligt hvor meget ekstra kvalitet man fik ud af at kræve, at der i øvrigt skulle være dit og dat og dut. Jeg synes peer review er rigtig centralt, specielt hvis man gør sig umage med at få reviewere ind over, som har indsigt i emnet. Så jeg har svært ved at se at man kan erstatte peer review med noget andet.

Question: *Jeg kom lige til at tænke på, nu kan jeg se at det ene tidsskrift her, det hedder Clinical Genetics, hvad hedder det andet? Modern Pathology. Det er også ovre i noget grundforskning, er det ikke?*

Response: Arh, Clinical Genetics burde være anvendt genetik, faktisk, clinical betyder anvendt. Det er rigtig svært at publicere sådan nogle weird emner som det her, og vi cykler rundt imellem tidsskrifterne, når den forrige har fået noget i et tidsskrift, så prøver man igen, og så gider de ikke have flere. Så vi publicerer i meget mystiske tidsskrifter. Eller ikke mystiske, men tit relativt kliniske, på trods af at det er noget ret nørdet, næsten grundvidenskabeligt vi publicerer.

Question: *Der er ikke et tidsskrift specifikt for?*

Response: Nej, det er for lille et emne.

Question: *Det var nemlig også det jeg lige kom til at tænke på, da du sagde I kun var fire. Så...*

Response: Arh, der er lidt flere, men af de sådan rigtige grupper, der er vi kun fire. Men det har jo med tidlig differentiering at gøre, så vi forsøger at publicere i tidsskrifter, der holder sig indenfor det emne, så har det med graviditet at gøre, så har det med cancer at gøre, så har det med genetik at gøre, med cellebiologi at gøre, så vi smugler os jo ind andre steder. Men man kan ikke bare holde et tidsskrift, og så bladre lidt i det og så holde sig orienteret. Vi publicerer alle mulige mystiske steder. Så det der man gjorde i gamle dage med at have journal clubs hvor man holdt øje med nogle bestemte tidsskrifter, det giver ingen mening mere. Her skal man PubMed-søge.

Question: *Ved du tilfældigvis hvor ofte de to artikler er citeret?*

Response: Nej. Den her (2) tror jeg er citeret ret tit.

Question: *Okay. Det var sådan set det jeg havde, medmindre du har mere at tilføje?*

Response: Nej, det tror jeg ikke.

A.2.4. Interview reference: M05

Date: Oct 27, 2011

Time: 15:45

Question: *Du har to artikler med, og dem har jeg jo bedt dig tage med, fordi du vurderede dem som værende høj forskningskvalitet. Og det vil jeg da meget gerne høre noget om hvad det er ved de artikler, som du vurderer som høj kvalitet.*

Response: Men det er for det første, at de rejser et, eller de beskæftiger sig med et væsentligt problem, som er af stor klinisk interesse. De gør det yderligere på en meget sober måde, her fordi de randomiserer til behandling og ikke-behandling - i det her tilfælde af mellemørebetændelse - om de skal have antibiotika eller ikke antibiotika. Dvs. de har flere grupper, nogen der får noget og nogen der ikke får noget, og de har randomiseret det. Det er sådan en meget grundlæggende god måde at gøre det på. Det er godt håndværk, forskningsmæssigt. Og så har de et stort materiale. De har næsten 300 børn med, så det er noget der batter noget, i stedet for et eller andet forkølet: "jeg prøvede det lige på 15 børn, og så skriver jeg om det, og så bliver jeg berømt." Den går ikke. Du skal gå i byen med noget substantielt. Og det er selvfølgelig også derfor, det kan komme i et tidsskrift som det her, for det er godt lavet det her. Så det er sådan set grundlæggende det jeg kan sige om det.

Question: *Er det bare den her (1), eller er det dem begge to?*

Response: Jamen det gælder sådan set lidt for dem begge to, og det er grundlæggende det samme problem de beskæftiger sig med. Den her (2) er så yderligere... det er sådan set igen behandling vs. ikke-behandling af akut mellemørebetændelse, og igen med placebo eller antibiotika. Og det er dobbeltblindet og randomiseret det her, så der er de yderligere ude i sådan noget med at sikre sig bias på resultaterne. Nu sidder jeg lige og kigger på - det her er mere end 300 børn, for der er 160 i den ene gruppe og 158 i den anden gruppe, så de gør det også ordentligt.

Question: *Så det er faktisk to studier der minder utrolig meget om hinanden?*

Response: Ja, de minder meget om hinanden. Men det er jo kliniske studier, som man kan sige vi er glade for, for det giver lidt mere fast grund under fødderne. Det her emne har været enormt diskuteret i vores kredse, så det er rart at der endelig kommer noget, der er substans i.

Question: *Ja.*

Response: Nå, men, ved du hvad... kan vi lade som om der er en tredje artikel?

Question: *Det kan vi sagtens, hvis du kan sende en reference til mig bagefter?*

Response: Det kan jeg godt. For noget af det som jeg godt vil lægge vægt på som en treer, for jeg har forsket en hel del metodeforskning. Altså sådan eksperimentelle metoder, eksperimentelle kliniske metoder, kan man sige.

Question: *Så det vil sige eksperimentelle behandlingsmetoder?*

Response: Nej, det er ikke behandlingen, det er simpelthen basale undersøgelser af fysiologien i mellemøret, hvis man kan sige det sådan. At det så kan have betydning

for behandlingen sidenhen, det er så noget andet. Det er jo klart at overordnet vil vi jo gerne have at i den anden ende gør vi det vi gør bedre for patienterne.

Question: *Det er målet?*

Response: Ja. Men jeg er enormt meget tilhænger af god, klinisk eksperimentel basalforskning, og det kan være f.eks. sådan at vi laver nogle undersøgelser, hvor vi måler mellemøretrykket, der findes forskellige metoder til at gøre det. Grunden til at mellemøretryk det er meget vigtigt at måle og kende noget om, og vide noget om hvordan det reguleres fra dag til dag og hos syge, det er at vi ser hos mange af vores syge ører, at de simpelthen har undertryk i øret, og så får de deformationer af trommehinden, og det laver nogle dybe lommer, og der sætter betændelse sig i, og så får de kronisk betændelse i øret, så får de erosion i deres øre, knoglekæden går i stykker, der kan ske forskellig anden ballade, hvor de får irreversible skader på hørelsen. Og det er så det vi som ørekirurg går ind og retter på nogle gange. Men der har man mange metoder til at måle det her tryk, og mest indirekte. Men problemet med det er, at hvis man måler det indirekte, dvs. udefra, så får man simpelthen fejl på målingen. Det man skal det er man skal bore hul, enten på knoglen bagved, eller på trommehinden, ind igennem trommehinden duer ikke, for så ødelægger man ligesom strukturen, så man skal måle det bagfra. Og det er, for at sige det mildt, møgbesværligt. Men vi har gjort det på nogen, der har haft en tumor i spytkirtlen, hvor man, når man skal ind og fjerne den, så åbner man ret bredt, for at passe på ansigtsnerven. Og der kan man faktisk få tippen af mastoid i feltet, hvor vi har fået lov til af etisk komité at bore hul og ligge et kateter ind og sætte en trykmåler på, og så sådan en sampledevice, ligesom når man måler EKG-døgnmålinger. Så monitorerer vi, og det kan vi gøre i et par dage. Vi kan også sætte dem i en stol og klippe slangen over, sætte en tregangshane på og ændre på trykket og lave eksperimenter med det. Og det har vi gjort. Og det er virkelig tungt, særdeles at rekruttere patienterne til det, dels fordi vi skal være et par ekstra mand til at lave det her, og i en hverdag at få det til at passe. Men vi får så mange gode, basale data ud af det, som ingen andre kommer med i virkeligheden. Så det her med at lave noget rigtig grundigt, hvor det måske nok kræver en helvedes planlægning, og det er tungt, organisatorisk at gå i gang med, men det giver nogle utrolig gode data. At man ligesom ikke hopper over hvor gærdet er lavest, men hvor man ser metodeproblemerne og alt mulig andet i øjnene og omgår de metodeproblemer, og gør det på den rigtige måde. Og i den sammenhæng også, for den sags skyld, at man tør koble flere ting på. Det er noget af det Asbjørn Drewes er berømt for, sammen med ham Hans Gregersen, jeg ved ikke om du er klar over det, men Hans interesserede sig jo meget for biomekanik, hvor Asbjørn han måler sansefysiologien, det er sådan nogle smertemålinger, men det de gør, de har et eller andet stimuli, et eller andet mekanisk stimuli, de udvider spiserøret, eller de udvider et stykke tarm, og så måler Asbjørn den sansefysiologiske respons; "gør det ondt, gør det ikke ondt, bliver det registreret med de der multikanals-EEG", og Hans Gregersen han måler i stedet for deformationen; hvor elastisk er det? Hvis du trykker på det med en bestemt kraft, hvor meget giver det sig? Så han måler elasticiteten. Så

ham her, han får noget ud om compliance, hvis man kan sige det sådan. Ham her, han får et eller andet neurologisk respons, plus han har også de der VAS-skalaer, så bliver man spurgt hvor ondt gør det, så skal man svare fra 1-10, så de har ligesom flere ting, og når de gør det på den måde, så kan de begynde at sammenligne, og det er også godt håndværk. Asbjørn har også fortalt, at somme tider har de sendt ting ind, hvor der er så meget, at de ikke engang kan publicere det i de der journals, for de synes det er for meget. Og det er det jeg egentlig gerne vil ind på når jeg laver de der forsøg med at måle på øret, for jeg har også lavet noget med Asbjørn med de her multikanals-EEG, hvor vi kobler det sammen, og så ser på, når vi laver de her trykændringer i øregangen, der giver trykændringer i mellemøret, hvordan det så bliver opfattet oppe i hjernen, hvad det er for nogle centre, der er involveret i - formentlig - at lave en eller anden overordnet regulering. For ligesom der er noget, der regulerer din hjerterytme, din vejrtrækning osv., så er der formentlig noget overordnet i hjernestammen, i hjernen, der regulerer trykket i mellemøret. Okay. Så det jeg er efter er gode, basale metoder, som ligesom kan arbejde med flere modaliteter, med flere eksperimentelle ting oveni hinanden, og også gern tværfagligt, for det her er der jo også ingeniører inde over. Så tværfaglig, basalforskning, hvor man virkelig skærer ind til benet, det har jeg meget stor respekt for. Og jeg vil gerne finde en artikel, der kan underbygge det her.

Question: *Lige et enkelt opklarende spørgsmål. Du sagde basal, klinisk forskning?*

Response: Ja, altså klinisk eksperimentel forskning.

Question: *Vil du ikke lige prøve at uddybe det lidt, hvad det egentlig betyder. For min opfattelse af klinisk forskning er egentlig i høj grad, at det er "ved patienten".*

Response: Ja, og det er også derfor jeg kalder det klinisk eksperimentel forskning, for det vi gør her er jo også ved forsøgspersoner, eller patienter, der indgår. Så derfor er det klinisk forskning, men det er eksperimentel forskning, for du udsætter dem for nogle eksperimenter. Klinisk forskning i sig selv, det kunne være at Sten Rasmussen derovre ville undersøge, om den ene hofteprotese er bedre end den anden, og så kører han hundrede af den ene slags, og hundrede af den anden slags, og hvis han gør det godt, så randomiserer han dem, og måske blinder han dem i forhold til røntgenlægen, der skal bedømme på et røntgenbillede, hvor godt de sidder, og sådan nogle ting. Det ville være klinisk forskning. Jeg kunne også godt lave klinisk forskning, hvis jeg ville undersøge om antibiotika det virker på mellemørebetændelse eller ej, men det her er klinisk eksperimentel forskning.

Question: *Og det er fordi I simpelthen udvikler en hypotese og så tester I om der er noget der, der kan virke, eller er det fordi I afprøver nye metoder, eller hvad gør det til eksperimentel?*

Response: Det er jo fordi man eksperimentelt... jaeh, her er det jo i virkeligheden flere ting. For her laver vi jo noget metodeudvikling, i det her jeg nævner med øret,

det er ikke lavet før, simpelthen. Så på den måde er det eksperimentelt. Men det er også eksperimentelt på den måde, at jeg så bagefter tager dem hen i en stol og klipper hul på slangen, sætter en tregangshane på, sætter en sprøjte på og så kan jeg hhv. lave undertryk og overtryk i systemet. Og så kan jeg på computeren, når jeg monitorerer trykændringen se, at et undertryk... hvordan modregulerer de det? Så der laver jeg jo også eksperimenter i eksperimentet, ik? Sten han vil jo ikke lave et eksperiment med hofteprotesen. Den sætter han jo i, kan man sige. Han laver måske et eksperiment på den måde, at han designer et klinisk forsøg, med den ene og den anden, men det er jo ikke eksperimentelt i den forstand.

Question: *Nej, jeg tror godt jeg kan følge dig. Du sagde noget andet, som jeg faktisk synes var rigtig interessant. I hvert fald interessant ordvalg; du snakkede om, at se det i øjnene. At se realiteterne i øjnene, tror jeg vist du sagde. Det var fordi du fortalte, at der var en undersøgelse, der var meget svær at lave, og det må man se i øjnene og gøre alligevel, og gøre det alligevel, fordi det er det, der er nødvendigt. Er det en problemstilling, tror du, I forskning generelt, at man ikke gør det? Eller er man god til det som forsker? At indse, hvad det er, der skal til? Eller er det ofte man ser, at folk springer over hvor gærdet er lavest?*

Response: Det tror jeg er meget, meget forskelligt. Fordi konjekturene de ændres, og jeg har i lang tid haft yngre kollegaer, der har fået deres stillinger i nakken, de er ikke interesserede i at lave forskning. Hvis nogen de var, så ville de måske gøre det, fordi de syntes det var sjovt. Dem har jeg så desværre ikke haft nogen af, men nu har vi et enormt tryk på vores speciale. Nu kommer de jo nærmest udefra, før de er ansat, og vil gerne lave projekter i afdelingen. Og mange af dem de er jo bare nervøse for deres egen fremtid. De vil bare være ørelæger, og synes det er florumvundet, og er frustreret over, at de ikke kan få en stilling uden de har skrevet en artikel eller tre, de kommer jo og er skideligeglade, de skal bare have noget papir. Og hvis det var jeg sagde vi ville lave nogle undersøgelser med mellemøretryk i en eller anden sammenhæng der, og sagde til dem, at vi bare kan gøre det med tympanometri, som jeg ved er en meget fejlbehæftet metode, så ville de med glæde æde den, de er da ikke interesserede i at gøre det mere besværligt for sig selv end højst nødvendig. De er interesserede i at få det gjort, fordi det skal gøres. Men det der er vigtigt, det er jo... der er et meget, meget velestimeret center i Pittsburg, der er nummer 1 i verden i forskning vedrørende mellemørebetændelse og det her undertryk i mellemøret. Og selv det center, de bruger den her tympanometri, selvom - jeg har været sammen med dem flere gange, og det er en af mine kæpheste det her, og de ryster på hovedet af mig hver gang. Jeg kan vise dem, at det er forkert, det de laver. De bruger de her til at undersøge de syge forhold i ørerne, og de analyserer deres data derudfra, og det duer ikke. Så selv sådan et center der, der er man dogmatisk i opfattelsen af tingene, eller bekvem kan man sige måske, så man ikke gør det man bør gøre. De er i øvrigt meget grundige med mange andre ting, kan man sige.

Question: *Men jeg synes det er rigtig interessant det du siger, og nu siger du det faktisk igen; de gør ikke det de bør gøre. Hvad så når man ser på forskningsartikler, sådan generelt, er det*

så det samme? Det lyder jo på dig som om det er stort set alle dem, der laver undersøgelser af mellemøret, der faktisk gør det forkert... eller problematisk.

Response: Nu er det jo ikke for at hænge alle ud. Og jeg skal forøvrigt også lige indskyde, at den der metode som jeg går lidt i rette med, den er udmærket til klinisk brug, altså til diagnose hos patienter. Så det er ikke fordi den ikke kan anvendes, men i en videnskabelig sammenhæng kan man ikke. Men det der selvfølgelig kan være problemet nogle gange, men det undskylder dem ikke, dem der i Pittsburg, det er kan være problemet er jo det muliges kunst. Og der er der nogle kompromiser nogle gange. En illustration af det er, at de der mellemøretryk, de er mest interessante at måle, der hvor det går galt. Dvs. de syge ører. Når jeg tager et bor hos en normal, der ikke har fejlet noget med øret og borer ind i området her, så allerede 1-2 mm inde, så er der sådan nogle luftfyldte hulrum, som kommunikerer op til mellemøret, så jeg sætter kateteret ind der, og så måler jeg hvad der sker. Men ser, du, hvis man har haft mellemørebetændelse, så er det her nede tit og ofte forkalket. Og så kan jeg ikke være sikker på, jeg kan bore ind og finde det her luftfyldte hulrum. Så kan jeg risikere at jeg skal til at bore endnu dybere, og den går ikke, for ansigtsnerven løber et eller andet sted tæt ved derinde bagved. Så det betyder at min begrænsning er, at jeg faktisk ikke kan bruge metoden hos de syge. Der hvor det ville være allermest spændende, der må jeg... jeg ved ikke hvordan jeg så skal gøre det... det må jeg gøre en anden gang, når jeg får en god idé. Så det er jo det muliges kunst, altid. Det er ikke fordi jeg vil sige, at... man kan jo ikke altid gøre det ideale. Det ideale skal nok være der, men du skal jo selvfølgelig være lydhør overfor hvad der er praktisk muligt. Men omkring det der, så synes jeg grundlæggende undersøgelser, hvor man går til biddet med metoderne, hvor man også meget gerne laver det multimodalt, og gerne tværfagligt, det synes jeg... det er tungt at danse med, men du får så meget mere ud af det. Det var derfor jeg var inde på det basale før, jeg ved ikke om man kan sige det her er basalforskning, men du får så meget ud af det, omkring de helt basale mekanismer.

Question: *Og i den forstand er det vel basalforskning.*

Response: Det er det et eller andet sted.

Question: *Også i den forstand, at det jo ikke er noget, der fører direkte til en ny behandlingsform. Der skal der jo så flere undersøgelser til, hvis det var.*

Response: Lige præcis. Det her er noget med nogle helt grundlæggende processer, og forstå dem, og så derfra bygge videre til kliniken.

Question: *Hvis vi nu lidt mere generelt... hvis jeg nu bare siger ordet forskningskvalitet, hvad tænker du så på når jeg siger det? Og det kan både være i forhold til artikler, eller i forhold til forskning i det hele taget.*

Response: Ja, det hænger jo sammen. Forskningskvaliteten, det handler jo om at man

har gjort sit håndværk. Og det mener jeg virkelig, for meget forskning det er jo håndværk, at kunne... dvs. at man laver noget ordentligt, substantielt, nyttigt, værdifuld og hæderlig forskning.

Question: *Har du lyst til at uddybe en lille smule på det, for det er jo nogle gode ord du bruger, men de kan måske fortolkes lidt forskelligt.*

Response: For at starte fra den anden ende - hæderlig, dvs. at man har integriteten i behold, at man står ved at beskrive det man finder.

Question: *Altså at det, der står i artiklen det er rent faktisk det man har gjort.*

Response: Det er sådan at man lige nøjagtig kan gå ind og reproducere det, hvis man vil det.

Question: *Så reproducerbarhed det er en form for kvalitet, eller en del af kvalitet.*

Response: Ja, det forudsætter jo at du er hudløst ærlig. Der er ikke smykket på noget - du beskriver i tilstrækkelige detaljer til at andre kan gøre det efter, og forhåbentlig finde det samme, som udtryk for at det er rigtigt, det du har gjort.

Question: *Så hellere beskrive nogle metodiske fejl man har begået, end at man prøver på at pynte på det.*

Response: Ja, eller i det hele taget, bare nogle dårlige resultater. Det kan også være nogle resultater, der forkaster din hypotese, din ellers så gode idé. Men i virkeligheden, det er sådan noget, det kan være svært at publicere, når man har negative resultater. Selvom det jo er lige gyldigt på den måde, at det også har værdi at beskrive for andre, hvordan man ikke skal gøre.

Question: *Det er det da. Og et eller andet sted bestemt i sundhedsvidenskaben. Der er jo nogle patienter det er gået ud over, i hvert fald i nogle tilfælde.*

Response: Ja, i nogle tilfælde. Så det er vigtigt, man skal ikke slette nogle af sine data, man skal ikke massere på sine data for at få en eller anden signifikans. Hvad sagde jeg så mere... det var noget med... Nyttigt, værdifuldt, det er jo hvis det er klinisk eksperimentelt, måske tenderende til noget basalt, så er det jo at det kan generere nogle hypoteser til klinikken. Eller hvis det er klinisk forskning, at det rent faktisk kan vise at den ene hofteprotese er bedre end den anden. Og i den sammenhæng er det ofte også vigtigt at man har gode observationstider på ting, at det ikke bare er noget man har undersøgt på to år, men at du har det på 20 år. Så det er jo også en kvalitetsparameter, at man har nogle ordentlige observationstider på tingene, eller på ens forsøg. For så kan man sige, at man opnår en behandlingsmæssig gevinst, som kan være god både for patienterne, men også for samfundet. De patienter skal måske ikke ind og have sat en

ny protese i, fordi de har fået en, der har fået en bedre holdbarhed, eller de her børn kommer sig hurtigere over deres mellemørebetændelse, får ikke et tilbagefald. Risikoen for at de så senere får kroniske følger er mindre. Så det var værdifuld.

Question: *Ja, jeg tror også primært det var de to ord jeg tænkte kunne betyde forskellige ting. Hvis man nu skal evaluere kvaliteten af forskning... først og fremmest; kan man egentlig det, og hvis du mener man kan det, hvordan?*

Response: Den er svær. Det er den. Den traditionelle måde er jo peer-review, men jeg ved ikke om jeg kan komme på noget, som er så meget bedre. Jeg er lidt tilhænger af det, de er begyndt på mange steder de senere år, med de her tidsskrifter med open access, eller online tidsskrifter, der er nogle stykker af dem, i hvert fald indenfor vores, fordi de er ikke så influerede af det her med at et tidsskrift har 100x12 sider om året, eller hvad de nu har af aftaler med deres publishers, det er en begrænsning, for så sidder de her reviewers der og i virkeligheden skal ikke kun bedømme kvaliteten, men skære ned på indholdet, og det er ikke sobert i virkeligheden. For du kommer til at sidde med nogle af de tidsskrifter, og sige nej tak til noget ordentlig forskning, til artikler, der i virkeligheden er meget gode, men som de bare ikke har plads til i det her tidsskrift, for de har ikke flere sider end de her, pr. år. Og det problem er du helt ude over ved de her online tidsskrifter, for de lægger det bare ind på en server, så der kan du gerne skrive to sider ekstra, eller du får ikke din artikler rejiceret pga. at de ikke har mere plads, og så skriver de i virkeligheden et review, som er lidt dodgy. Altså at de kritiserer nogle ting, der i virkeligheden ikke er så væsentlige, for at sige, at vi ikke kan publicere det her. Så det der review, jeg tror egentlig grundlæggende det er meget godt, men det skal dermed sagt ikke være begrænset af en eller anden kontrakt om et antal sider, som en journal har med et forlag.

Question: *Er det erfaring du taler af, eller er det mere noget du har læst om/hørt om?*

Response: Jeg ved ikke om jeg har haft direkte erfaring med det - det har jeg måske uden jeg har vidst det. Fordi det er jo svært nogle gange at få tingene publiceret... men der jo nogle gange, det der med at skrive og få publiceret, det tager jo mindst lige så lang tid som at lave forsøgene nogle gange.

Question: *Men nu vi snakker om open access og online tidsskrifter, kender du til det begreb, der hedder open review?*

Response: Nej.

Question: *Det er heller ikke fordi det er så vigtigt, jeg kom bare til at tænke på i den her sammenhæng; der er nogle få tidsskrifter, der er begyndt på at gøre det, hvor den traditionelle proces bliver erstattet af, at når du sender din artikel ind, så bliver den præpubliceret, på nettet, hvor alle i princippet kan gå ind og lave et review af den, men hvor de så også kommer til at fremstå med navns nævnelse, så alle kan også se hvem der har reviewet, og*

hvad de har skrevet. Det giver jo en større åbenhed, og måske også en større ærlighed. Har du en umiddelbar reaktion på den proces, når du hører om den?

Response: Det kunne måske egentlig være udemærket, men det forekommer mig også lidt uoverskueligt, for du kunne godt forestille dig, at det kunne blive temmelig uensartet. Og hvornår er det så lige, der er nogen, altså, hvordan garanterer man, at der er nogen der reviewer den artikel? Når man sender den ind til en journal, så bliver den sendt til en editor, der sender den til en sub-editor, der er inde i det område, der så sender den til nogle reviewers - det er ligesom organiseret. Det andet forekommer mig lidt uorganiseret. Du skal forlade dig på, at nogen har en litteraturservice, hvor den der den dukker op, den er præpubliceret, så går de ind og kigger, og så er de villige til at gå ind og bruge tiden på at lave et review. Når vi andre bruger tid på at lave review, så er det jo fordi vi får en direkte forespørgsel fra editor, om ikke vi vil bruge tid på at kigge på en artikel. Og så føler man jo en eller anden forpligtigelse til at gøre det, for man er jo en del af det der, også når man selv sender artikler ind. Men hvad var det du spurgte om, et godt tidsskrift? En god bedømmelse?

Question: *Nej, egentlig bare din reaktion på det, på hele idéen.*

Response: På open review... summerende så vil jeg tro, at det kunne blive rimelig uensartet og ustabil. Hvem er det, der har tid til at gøre det, hvem er det, der gør det? Hvis ikke lige de bliver bedt om det?

Question: *Det kunne man i hvert fald være lidt bekymret om.*

Response: Ja, det vil jeg sige, det er næsten nogen, der har for god tid., eller som er nogle kværunter.

Question: *Det er jo ikke til at vide. Det var egentlig heller ikke så vigtigt, men jeg kom lige til at tænke på det, og syntes det var interessant. Men, nu nævner du jo faktisk ret tidligt her, at det er randomiserede, dobbeltblindede, kontrollerede, prospektive studier - det har jo også noget med evidensniveau at gøre. Er der et forhold imellem evidensniveau og forskningskvalitet?*

Response: Jaeh, det er der vel. Når jeg svarer lidt undvigende, så er det fordi, det der med evidensniveauer, jeg forstår godt, hvad du mener med det, men jeg er faktisk ikke helt klar over, hvordan man definerer de forskellige grupper. Og det har noget at gøre med, at det er et begreb, der er kommet ind i forskningen, som er af nyere dato, end da jeg gik på ph.d.-kurser. Så jeg kender det, men meget af det forskning jeg laver, det er jo ikke sådan klinisk, medicinsk forskning, hvor man randomiserer til den ene og den anden behandling. Mit, det er mere klinisk opgørelse af behandlinger og så de der klinisk eksperimentelle ting. For det du snakker om, det er jo sådan noget, man kan bruge til at karakterisere sådan noget, som det her med. Hvor man randomiserer en behandling, eller også gør man ikke, men man har to grupper, som man sammenligner

med, og så er niveauet ikke så højt, som hvis man randomiserer.

Question: *Ja, eller man har bare et par cases.*

Response: Ja, præcis, hvad hedder det så?

Question: *Jeg kan ikke lige huske hvad de hedder, og der er lidt forskellige inddelinger, alt efter hvem der har lavet dem. jeg tror der findes en tre-fire forskellige skalaer, men generelt så ligger de - i hvert fald i forhold til behandling - ret langt nede i hierarkiet. Men så tænker jeg på - er der så forskel på hvor vigtigt du vil vurdere det, at det er et dobbeltblindet forsøg, afhængig om du har dine kliniske briller på eller dine forskerbriller på?*

Response: Nej. Der synes jeg, der er jeg så meget forsker OG kliniker, det kan jeg ikke skille ad. Der synes jeg man må holde fanfen højt, også selvom man er kliniker, og ligesom vide, at der er altså en golden standard, som man må sige... man kan selvfølgelig diskutere nogle gange, ikke, hvor omfangsrigt det skal være, hvor højt niveauet skal sættes. Men sådan noget, f.eks., som at lave noget prospektivt og randomiseret, det er så basalt, at det bør man altså gøre.

Question: *Jeg tror egentlig også mere jeg tænkte den anden vej rundt. At som kliniker havde man brug for de her store, randomiserede, kontrollerede forsøg med mange patienter, hvorimod som forsker, der kunne man måske i højere grad også brug cases, in vivo og in vitro forskning, og hvad der nu ellers er af forskellige studietyper, til simpelthen at være hypotesegenererende, idégenererende...*

Response: Det kan man sagtens, det er jeg meget enig i, ja.

Question: *Så som forsker er der måske lidt større frihed i hvad man måske kan bruge til noget, end som kliniker? Eller kunne man som kliniker måske også bruge en kasuistik?*

Response: En kliniker kan jo sagtens bruge en kasuistik og så gå i byen med det.

Question: *Kan de tillade sig at omsætte det til en behandling?*

Response: Nej. Men du kan jo godt få en idé og så måske - hvis du nu siger, det skulle være case-baseret, så skulle man jo nok yderligere underbygge det med noget litteratur, hvis der er en idé med et eller andet, til at generere en hypotese om, at en eller anden behandling er god for hvad nu- Så kunne man jo godt derfra gå i byen derfra og lave en pilot, som kliniker, og så derfra bevæge sig videre til en egentlig prospektiv afprøvning.

Question: *Men så er vi jo så faktisk også ude i, at det var forskning man lavede på baggrund af det, og ikke en klinisk behandling.*

Response: Ja.

Question: *I hvert fald ikke i første omgang. Okay. Så har jeg ikke så meget tilbage. Det sidste jeg har er en brainstorm, hvor jeg har nogle begreber, jeg gerne vil høre din umiddelbare reaktion på, i forhold til forskningskvalitet, og mange af dem har du faktisk allerede svaret på, men det første jeg har er "ny viden".*

Response: Jamen ny viden det er positivt, det er godt, det skal vi eftertrage, for ny viden det giver os mulighed for at gøre tingene måske lettere, måske bedre for os selv og patienterne, måske mere økonomisk, måske mere miljørigtigt, besparende. Ny viden er simpelthen kilden til fremgang.

Question: *Ja, "layout af artikler"?*

Response: Det betyder utrolig meget. Det synes jeg, fordi... jeg er sådan lidt freak med layout af tingene. For hvis noget er skrevet sjusket, så synes jeg næsten det kan være lidt en overvindelse at læse det. Tingene må meget gerne være lækre. Det gør jeg selv meget ud af, at formulere ting pænt, men også sætte det pænt op.

Question: *Så både sprogligt og rent visuelt?*

Response: Ja.

Question: *Betyder det noget for kvaliteten?*

Response: Nærjh, det gør det ikke nødvendigvis, men jeg vil sige, jeg ville have mine forbehold, for hvis folk ikke kan finde ud af at stave ordentligt, og hvis de ikke kan finde ud af at lave en tabel, så den er overskuelig, hvordan skulle de så kunne finde ud af at holde styr på deres data? Og det kan godt være de kan det, men det giver lidt en... det kan godt give anledning til en vis reservation, synes jeg. Altså, jeg har haft diskussionen mange gange med mine yngre kollegaer, de kommer med et manuskript, hvor de synes de har gjort et vældig stort stykke arbejde, og skrevet noget på engelsk og siddet og svedt over det - så kan de ikke engang finde ud af at sætte en stavekontrol på det. Hvad er det for noget? Vi snakker om højtuddannede akademikere, så skal jeg sidde og kigge på stavefejl? Jeg bliver simpelthen sur, jeg gør... nå, men på den måde så giver det sådan en (rejective sigh), og et dokument jeg får, nu snakker vi måske på manuskriptplan, som vi sender imellem os, hvor den ene del af dokumentet er skrevet med lige margen, og den anden del er skrevet med venstrejusteret margen... altså... hold nu op, lav det ens. Jeg synes godt det kan - ihvert fald på det niveau - give anledning til lidt frustration og reservation i forhold til de pågældende kollegaer. For jeg føler, at hvis de har orden i de ting jeg kan se, så har jeg også større tillid til, at de har orden i de andre ting. Og det er i virkeligheden lidt det samme når man går videre. Nu er der jo nogle professionelle editors, der sørger for at tingene som regel ser pænt ud, og følger et bestemt design, nogle er moderne, nogle mere klassiske. Men dybest set, så er der jo som regel nogle ting som går igen, som skaber en god oversigt; kapitlerne, abstract... og det... jeg kan ikke huske jeg har set en artikel, som jeg syntes var sjusket i layout. Jeg er ikke stødt på

det - men jeg ville falde over det, hvis jeg så det.

Question: *Hvad med sådan noget som impact?*

Response: Det er jo lidt det samme som det der før med review... har vi noget der er bedre? Jeg kommer jo fra et af de små specialer, og hvis jeg publicerer i det højest rangerende øre-næse-halstidsskrift, så kan jeg måske svinge mig op på en impact factor på 2, små 3 måske. SÅ jeg synes ikke impact factor er en skidegod idé. Min kollega Jeppe Hagstrup nede fra nefrologisk afdeling, han skriver selvfølgelig i et nefrologisk tidsskrift, de er jo helt anderledes store og brede, og har jo lynhurtigt en impact factor på 8 eller 10. Så...

Question: *Men vi kan jo se, du kan jo også komme i New England Journal of Medicine.*

Response: Ja ja, joeh, men det vil jeg sige, det er meget sjældent sådan noget. Men det er rigtigt, det kunne man godt. Hvad har de? tyve?

Question: *Nogle-og-halvtreds, så vidt jeg husker. De er rigtig højt oppe.*

Response: Men, ja, jeg ved ikke hvad jeg skal svare, for jeg har svært ved at komme med noget, der er bedre. Jeg synes man skal i hvert fald vide, at det er meget afhængigt af ens fag, hvad område impact factoren ligger i. jeg har søgt nogle penge ved forskningsrådet, og de går meget op i impact factor, og i den sammenhæng synes jeg det er dybt uretfærdigt, for jeg kan adrig nogensinde komme i nærheden af de andre, selvom jeg publicerer virkelig god forskning. Det kan jeg bare ikke.

Question: *Hvad så med citationer?*

Response: Jaeh. Jeg er ikke helt sikker på det, men jeg vil da tro, at det samme forhold vil afspejle sig, at det også vil være meget fagspecifikt, hvor niveauet ligger henne. Men jeg vil da gerne have, at nogle af mine afgørende, basale forsøg med mellemøretrykket de bliver citeret af rigtig mange; det tror jeg i grunden også de bliver. Og dermed kan man sige, også blive regnet som væsentlige artikler. Men jeg vil formentlig stå med, at der vil ikke være så mange citationer på det, som med en sammenlignelig artikel i endocrinology eller hvad de må hedde i de store medicinske.

Question: *Men hvis du nu skulle vælge imellem, at dine artikler de blev evalueret på baggrund af enten impact factor eller citationer, hvad ville du foretrække?*

Response: Jeg tror da egentlig jeg ville foretrække citationerne.

Question: *Hvorfor tænker du det?*

Response: Jo fordi det forekommer mig mere ærligt. Fordi impact factoren - jeg kan jo ikke være sikker på... jeg synes egentlig noget af det forskning jeg har lavet, det

fortjener egentlig en bedre impact factor, i hvert fald det eksperimentelle af det, ikke nødvendigvis de kliniske ting, men jeg har haft svært ved at få det publiceret hvor... jeg ved ikke om det er generelt, men jeg er jo heller ikke så god til at skrive engelsk, som når man er englænder eller amerikaner. Det flyder jo lidt lettere når man er indfødt - men jeg tænker lidt, at der hvor jeg ville få glæde af en høj impact factor, det er hvis jeg fik en artikel, der måske ikke helt fortjener det, i et tidsskrift med en høj impact factor. Men jeg synes snarere det modsatte er tilfældet. At jeg får nogle gode artikler i nogle tidsskrifter, hvor jeg egentlig satser højere, men så er nødt til at gå lidt ned og så få det publiceret i et tidsskrift, der måske er knap så restriktivt.

Question: *Så citationerne afspejler i højere grad virkeligheden end impact faktoren?*

Response: Ja, og de afspejler jo også interessen for min artikel. For impact faktoren gælder jo for en journal. Og så ved jeg godt, det er baseret på de enkelte referencer, men citationen er jo lidt mere min. Er det ikke rigtigt forstået også? For det er på den enkelte artikel, ja?

Question: *Jo. Og de andre ting jeg har, dem har du sådan set svaret på, så jeg tror ikke jeg behøver at spørge dig om tværfaglighed, det har du giver meget tydeligt udtryk for. Så jeg vil sige tak.*

A.2.5. Interview reference: M08

Date: Oct 17, 2011

Time: 14:30

Question: *Du må meget gerne fortælle mig noget om de her artikler, selvfølgelig med fokus på hvad det er ved dem, som du opfatter som høj kvalitet.*

Response: Den første artikel, den er som sagt publiceret af vores egen forskningsgruppe, og grunden til at jeg opfatter den som høj kvalitet er, at den tager fat i en klinisk problemstilling, anvender nogle registre vi har tilgængelige, som er ret unikke for Danmark og kommer til en konklusion som får en stor betydning for patienter, der har den sygdom der hedder Barrett's esophagus, nemlig det at de ikke får kræft alligevel. Og det vigtigste i den her artikel... en måde at illustrere det på, det er den her figur som hedder 1, hvor du ser på hvor hyppigt man fik kræft i spiserøret, det her er summary, den er lidt fræk den figur, for nu siger vi at vi tager alle de tidligere uden at lave noget rigtig videnskab over det, lægger dem alle sammen og puljer dem, og så kigger vi på vores, afhængig af om det er drenge eller piger eller hvad det er, og så ligger de helt dernede. Figuren er fræk fordi man har taget alle de andre uden det er særlig videnskabeligt og så fået den i *New England Journal of Medicine* alligevel, og så sagt sådan gør vi det. Så er den selvfølgelig understøttet af nogle andre figurer og noget. Og det jeg synes der er unikt ved den her, det er mere det at den tager fat i noget der er utrolig vigtigt og an-

vender noget der er ret unikt for Danmark i og med man har sådan nogle gode registre. Og det synes jeg egentlig er ret godt, og det synes de så også i *New England Journal of Medicine*.

Question: *Så det er en kombination af det design I har brugt og så at det er en klinisk relevant problemstilling?*

Response: Jeg synes ikke designet er særlig originalt, det er sådan man gør, det er sådan et case-control studie.

Question: *Så det er de data I har?*

Response: Det er vores data, og så idéen, at man er skeptisk over for om det her det er rigtigt og så går man tilbage og finder alle de patienter. Men man kan sige videnskabeligt er det en slavisk måde at gøre tingene på, det er mere at man får den der idé, der får en hel masse impact. Og så er metoden state-of-the-art. Der er ikke noget nyudvikling over metoden, det er bare en god, robust artikel.

Question: *Hvad med den næste her, nummer to?*

Response: Den næste har noget at gøre med et nyt lægemiddel, der hedder pregabalin, som er et smertestillende middel, som behandler patienter med kronisk pancreatitis. Og grunden til at den er god, det er at den er resultat af ti års forskning. For det første har vi, baseret på patienter vi har inde, hvor vi har lavet måske ti artikler, de har en speciel type smerter, nemlig neurogene smerter, ligesom når man får amputeret en arm, så kan man få amputationssmerter. Og på samme måde har det vist sig, at patienter der har fået ødelagt deres bugspytkirtel, de kan få ødelagt nerverne og dermed få neurogene smerter. Baseret på alle de studier vi har lavet, som egentlig nogle andre har lavet histologisk, vi er de eneste der har lavet noget neurofysiologisk på det, så tænkte jeg hvis man skal gøre det, så skal man bruge det lægemiddel som man bruger til behandling af neurogen smerte til at behandle de her patienter, og det skal siges at de næsten alle har smerter, og det er det de dør af, for de kan ikke spise, så ondt har de. Så holder de op med at spise, så dør de som muselmænd, åreforkalkning, alt sådan noget. Så sætter vi det her studie op, og det gjorde vi en del år tilbage med forgængerer af det stof der hed pregabalin, det hed gabapantin, undervejs udløb Pfizers patent på det, så var de ikke særlig interesserede i at støtte den undersøgelse, kun med det der hedder pregabalin, som er noget af det samme stof, det har bare en bedre kinetik. Og så kom vi igang med det, og det vi gjorde var at vi kontaktede Pfizers hovedkvarter i England, der lå det dengang, og vi fik en snak igang og en dialog omkring at vi havde de her patienter, at vi gerne ville afprøve det, vi fik også hugget et grant fra dem, faktisk på to millioner kroner, så kumulerede vi en international gruppe, fordi vi ved vi har en gruppe i Holland som har mange af de her patienter, som vi kendte fra tidligere, så etablerede vi et samarbejde med dem og fik studiet igennem både i Danmark og i Holland. Så gennemførte vi det, fandt ud af det faktisk havde en

effekt og publicerede det i den bedste journal som vi kan få det i, nemlig *Gastroenterology* som har højest impact i gastroenterologien. Og man kan så spørge hvorfor jeg synes det er særlig unikt, for studiet er egentlig også state-of-the-art. Det der er unikt ved det er for det første, at man laver et multicenterstudie her fra Aalborg uden det er et firma der går ind og overtager styringen, langt de fleste multicenterstudier er nogen som er firmainitierede, for de har netværket og mulighederne for at lave det. Det er kolossalt svært at lave, altså få det til at fungere i de andre lande, få dem til at inkludere patienterne, altså hele den der proces, den er meget vanskeligere end man tror, altså at få det til at fungere i to forskellige lande. Og så også at der ligger så meget basal forskning op til det her, at man siger, "baseret på det vi tidligere har lavet, så laver vi det her studie med lægemidlet, som skal virke på specielt den type smerter vi har", og så virker det rent faktisk også. Men det kunne også være det ikke havde virket, men hele idégrundlaget op til, og så laver man et studie der kører efter alle retninglinier, og får det publiceret, men også det at man rent faktisk etablerer det netværk og får den anden gruppe med til at lave det, det var svært kan jeg sige dig. Det har været meget svært, selvom det er nogen vi kender i forvejen, altså i Holland fungerer tingene helt anderledes, selvom det er EU-lande, Holland og Danmark. Etisk komité er professionel og alt sådan noget. Der er mange svære ting. Så det er derfor jeg synes den er god.

Question: *Og den sidste?*

Response: Den sidste er publiceret i *Pain*, som jo er flagskibet indenfor smerteforskning, men ikke en journal der har impact på højere end en 4-5 stykker, men det er jo også fint. Der er vi meget glade, hvis vi får noget publiceret. Og grunden til at jeg har taget den frem er, at den er lavet af forskere her fra Aalborg og så fra Bern, og man er gået sammen med nogle statistikere, for man har en meget vanskelig problemstilling at skulle løse, altså man har kigget på om der var sammenhæng imellem forskellige mål for smerte, altså eksperimentelle mål, om der er nogen sammenhæng. F.eks. om du er overfølsom over for varme, eller om du er overfølsom overfor kulde, eller varme og strøm, og man har altid troet der var den sammenhæng, det er der også mange der har vist, men her er man gået ind og har lavet en faktoranalyse, og fået meget tung statistik lagt ind over, hvor man ligesom har fået meget hjælp, og det er et kæmpestort materiale, jeg tror det er 300 personer, der har fået lavet alle mulige mål, og så er man gået ind og set at det har fordelt sig på nogle forskellige clusters, og de clusters er totalt uafhængige, det er ligesom et stort haglbøsseskud, når man ser på sammenhængen imellem dem. Og det er vigtigt, for vi har egentlig altid troet, at hvis man bare málte een smertemodalitet, f.eks. varme, så vil den sige noget om smertesystemet, f.eks. om det virkede at give morfin. Og der var overhovedet ingen sammenhæng imellem de forskellige smertemodaliteter, det passer hånd i hanke med det vi altid har undret os over og også fundet. Jeg synes det er flot arbejde de har lavet, de har faktisk lavet rigtig stort materiale, de har lavet rigtig mange smertemål, det er sindssygt stort. Det er på tværs af forskellige forskningsgrupper, og så har de også etableret et samarbejde med en ny gruppe, nemlig nogle statistikere, for at få hele det der til at svinge.

Question: *Og det er noget helt nyt de har fundet ud af?*

Response: Nej, man har før diskuteret om de var afhængige, men mange af de andre har fundet at de var afhængige, og det er også nemmere at få publiceret, men nu har de bare lavet et studie, der minder lidt om vores med Barrett's esophagus, hvor vores materiale var meget større end de andres, så får man ikke alle de selektionsbias, som på de små materialer. Det har man så gjort her, og så finder man nada - og det er vigtigt, for så kan vi sige varmetærsklen er uafhængig af kuldetærsklen, så hvis vi skal arbejde med smertesystemet, så må vi have mange forskellige test for at efterligne det bedst mulig.

Question: *Så både nr. 1 og nr. 3, det er sådan set et spørgsmål om at man har kunne afvise noget tidligere forskning som værende besluttet forkert?*

Response: Arh, ja, de har jo altid gået lidt op og ned, men i den første artikel har langt det meste tydet på, at de her patienter kunne få kræft i spiserøret, og da det er 1% af befolkningen der har det, det har jo en sindssyg høj prævalens, det der Barrett's esophagus, så er der jo vildt mange mennesker der er gået rundt i screeningsprogrammer, som er unødvendige. Men der var altså studier der har vist også, at det nok ikke var så nødvendigt, bla.a. et tidligere dansk. Men vores viser at sandsynligheden for at få kræft, den er så lille, udover det første år, at det slet ikke kan betale sig at kontrollere dem. Så det kan man godt sige afviser de andre. Med så meget overbevisning at jeg tror det får en stor betydning for klinikken. Det her (3) er så ovre i det eksperimentelle, og det afviser også hvad andre har fundet, langt de fleste har fundet en sammenhæng, men det er meget vigtigt for os, for at forstå basalforskningen, grundforskningen. Og nr. (2), der kommer vi med en ny hypotese.

Question: *Du nævner jo faktisk rigtig mange forskellige ting her som kvaliteter. På den ene side siger du det har betydning i klinikken, det går jeg ud fra er en kvalitet ved et studie, eller det kan det være?*

Response: Ja, den er svær. Vi ved jo godt, at det der fører til kliniske ændringer det er jo basal videnskab, det der langt hyppigst fører til kliniske ændringer, det er basal videnskab, der intet har med klinikken at gøre. Her kan vi bare se den røde tråd (refers to paper), her kan vi ikke se nogen rød tråd, men den kan komme sidenhen (refers to paper). Hvis du spørger... jeg var engang til et eller andet forskningsseminar, sammen med Rehfeldt tror jeg nok, han holdt et foredrag og viste en tabel, hvor man havde spurgt forskellige klinikere hvilken artikel de synes havde været epokegørende for ændring af deres praksis. Så pegede de næsten allesammen på en eller anden basal artikel og ikke en klinisk artikel.

Question: *Og med basal da mener du grundforskning?*

Response: Ja, det gør jeg. Jeg mener stadig at grundforskning er det, der flytter noget.

Question: *Jeg synes det er rigtig interessant at du siger det...*

Response: ...men jeg synes det er en god artikel, for et eller andet sted skal du bringe det helt op til klinikken. Men det er tit og ofte at klinikerne har svært ved at forstå hvad vi laver alt det der basale for. Faktisk er det det, der flytter bjerget, tit.

Question: *Jeg synes det er rigtig interessant, for nu har jeg snakket med en del andre, og en af de ting mange har peget på som kvalitet, det er at det er en klinisk relevant problemstilling, fordi det så er noget de kan bruge i deres hverdag. Og ud fra det du siger, så vil jeg jo umiddelbart mene, at nr. (1), den har en klinisk relevant problemstilling, hvorimod nr. (3) måske ligger lidt længere væk, netop fordi det er grundforskning.*

Response: Jamen det er fordi du stiller et spørgsmål jeg har kolossalt svært ved at finde... det var derfor jeg sagde til dig, "Hvilke tre artikler?", hvis du havde spurgt mig, "for at gå stuegang oppe på afdelingen", de her tre havde været helt uinteressante.

Question: *Jamen det er jo interessant for mig uanset, for hvad er forskningskvalitet?*

Response: Jamen vi har jo mange kasketter på, jeg kan jo komme op og være den klinisk ansvarlige på afdelingen, så vil jeg sige det her (1), måske det her (2) og slet ikke det her (3), men skal jeg sidde som forsker og udtale mig, så bliver det den her (3) der er nr. 1 i rangfølgen fordi den basale forskning er det, der flytter klinisk forskning. Men det kan være der går ti år.

Question: *Så det du siger er, at for at den der er klinisk relevant, den der er tæt på den kliniske hverdag, den overhovedet kan komme til, så skal grundforskningen være på plads først?*

Response: Det er lige præcis det, det er grundforskningen hvor de store landevindinger er, det er der ingen tvivl om, det vil alle være enige i, også dem du har spurgt, hvis de ellers får præsenteret spørgsmålet rigtigt. Men derfor kan det jo godt være den kliniske hverdag vi forholder os til, det er den om der er en øget stigning af kræft, eller om pregabalin virker på kronisk pancreatitis. Men jeg mener faktisk, selvom jeg har taget to kliniske artikler med, at grundforskning er det vigtigste. Men det er igen, det kommer an på hvordan spørgsmålet stilles.

Question: *Hvad kommer med på referencelisterne i guidelines? Hvis du skulle skrive en ny klinisk retningslinie, hvilke af de her tre kommer så med?*

Response: Det gør de her to (1 & 2).

Question: *Der er en af mine deltagere der har nævnt, at der er en slags fødekæde. Vil du også betegne det på den måde? Altså man kan sige at grundforskningen ligger i bunden af fødekæden, de bliver nødt til at være der for at man kan bygge ovenpå.*

Response: Joeh, det er da den ideelle situation, det er bottom-up. Men den findes ikke, fordi langt den vigtigste videnskabelige observation, det er jo case-studiet, det er tilfældighederne. Man gør et eller andet, og så baseret på et eller andet tilfældigt noget, så går man tilbage og finder at det nok passer med den grundforskning man havde, og så går man tilbage til klinikken igen. Det er jo det, der sker i realiteten. Case-studiet det er faktisk, selvom det ikke er videnskab, langt det vigtigste til at gøre nye opdagelser. Men så gør man det, man finder ud af tilfældigt, at et eller andet medicin virker på en eller anden lidelse, så dykker man ned i grundforskningen og siger, "Hov, der er faktisk belæg for at det faktisk er rigtigt", og så går man op til klinikken igen. Så det er sådan et V, og det er der, de store opdagelser er. Den ideelle situation er at grundforskningen fører frem til kliniske ændringer, og jeg kender kun til eet eksempel, og det er indenfor HIV-behandling, hvor man rent faktisk besluttede at man ville lave nogle nye protease-hæmmere, altså som virker imod virus, og man satte sig ned og arbejdede på det, og så fandt man også frem til noget. Så fik vi så et spin-off i leverbetændelse, for vi kunne bruge det lægemiddel til leverbetændelse også, og det er igen sådan en case-story mere eller mindre.

Question: *Så den her fødekæde idé det er mere sådan en slags ideel verden, som ikke helt findes?*

Response: Jeg kender ikke andre end HIV-eksemplet, hvor det rent faktisk lykkedes at arbejde seriøst på den problemstilling, hvor grundforskningen fra start af havde båret op. Det er hyppigt at grundforskningen går i lidt forskellige retninger, så fører den til en eller anden idé, og så går den op til klinikken. Altså, der er jo nok en vis form for fødekæde, det er jo ikke fordi den ikke er der, men det er jo slet ikke så stringent som man går og bilder sig ind. Det er jo tit det er post-hoc at man går ind og siger, "Nå ja, det virkede, så går vi til grundforskningen, ser om det passer, og så går vi op igen", så det er sådan jeg synes det rent faktisk fungerer. Men stadig mener jeg jo at vægten skal lægges i grundforskningen, det er der de store fremskridt er. Det andet pudrer det bare lidt. Det er fordi man har en klinisk synsvinkel på det, man kender grundforskningen, og så siger man, "Nåh ja", altså det vil sige man hopper imellem de to lag hele tiden. De hænger ikke kontinuert sammen.

Question: *Hvis vi så prøver at snakke lidt mere generelt end lige de her artikler, og så forskningskvalitet helt generelt, hvad tænker du så på?*

Response: Det er et svært spørgsmål at svare på, og det er jo derfor du spørger. Jeg ser f.eks. ikke sådan på det, at den her artikel (1) er den bedste, fordi den har en impact factor på 53, det gør jeg slet ikke. Det kan være at den her artikel (3) med den laveste impact factor er den bedste, fordi det er den jeg måske synes udstråler den største kvalitet, og hvordan tingene er gjort rationelt. Jeg vil sige forskningskvalitet for mig, handler mere om hvordan hele processen er kørt fra start af, fra idéfasen til protokolskrivning til studiet er gennemført. Hvordan er samarbejdet med andre? Hvordan er artiklen

blevet? Hvis man så er uheldig at den ikke viser noget som helst, så får man det ikke i de fine tidsskrifter. Det har ikke noget med forskningskvalitet at gøre, eller negative resultater, kvaliteten kan være mindst ligeså god, eller er formentlig mindst lige så god, men mange af de ting kommer jo aldrig til dagens lys, for tidsskrifterne selekterer hårdt på at banke deres impact factor op, og de negative studier de giver ingen impact, de bliver aldrig citeret. Så det skal man i hvert fald passe på med. Men så kan man så sige, hvad er så forskningskvalitet? Hvis vi nu sagde det handlede om at bygge hele processen op til at få den færdige publikation ud, er det så nødvendigvis forskningskvalitet, at det er fulgt stringent? Det synes jeg er svært at sige, for kvalitative ting er per definition svære at måle, og det kunne ligeså godt være, at man har haft en fantastisk proces, med et eller andet resultat, men det allervigtigste måske, er at f.eks. en ph.d.-studerende pga. den proces har udviklet sig til en dygtig forsker, der sidenhen bar forskningen videre, så det kunne være at vores kvalitet i virkeligheden lå et helt andet sted end selve det produkt der kom ud, det kan også være at der slet ikke kom noget produkt ud, men at processen har været god og tømmer gruppen sammen. Så derfor er det ikke entydigt at svare på det her.

Question: *bestemt ikke, og det forventer jeg jo heller ikke.*

Response: Men det er jo tit spin-off, hvordan er miljøet i gruppen, hvad bærer det mere med sig, hvilket kontakter har vi fået? Jeg tror det er vigtigt at skabe de rigtige mennesker, altså, gå ind og kig på min forskningsgruppe derinde, og se hvem det er der sidder derinde, er det nogen der er dynamiske? Er det nogen der efter deres ph.d. også bliver ved med at være forskere? Sådan noget det er forskningskvalitet hvis de er det.

Question: *Så sådan noget med at skabe forskningsnetværk i forskningen, som led i forskningsprocessen, at skabe netværk, det er også en del af det?*

Response: Det er da bestemt vigtigt, men det er ikke kun netværk, det skal man ikke lægge for meget vægt på. Vi har et kæmpe netværk, vi har alle de her samarbejdslaboratorier rundt omkring i verden, men det rigtig gode forskning vi laver det foregår altså herhjemme i Aalborg. Vi bruger utrolig meget energi på at skabe netværk. Men hardcore forskning, der virkelig flytter noget, og ændrer ting, det foregår her i Aalborg, eller i Århus, det er jo det samme. Men det foregår i vores lokale gruppe, det gør det.

Question: *Vi snakkede lige kort om sådan noget med design, noget med at det ene var noget case-control, og det hænger jo også lidt sammen med de her evidens-niveauer.*

Response: Ja, selvfølgelig, det kan det let gøre.

Question: *Er der en sammenhang imellem det og så kvalitet?*

Response: Bestemt da, for evidens-niveau A, det kræver at der ligger en god, kontrolleret klinisk undersøgelse, eller flere, til grund for det. Så det vil sige, det at under-

søgelsen gennemføres korrekt, som case-control, som kontrolleret, klinisk undersøgelse, dobbeltblind osv. Det vil kunne ligge til grund for niveau A, hvorimod en eller anden kasuistik har et niveau der ligger nede på C, eller hvad det nu er vi plejer at rate dem som, eller D.

Question: *Så der er en eller anden sammenhæng?*

Response: Jamen det er der da bestemt. Det er da sådan nogle af dem som bliver citeret og som bliver brugt til en evidensniveau, men man skal også passe på der, Jens Peter, at lægge for meget ind vægt i det evidensniveau. Hvis du går op og kigger på vores afdeling og hvad vi laver til hverdag, så er det altså ikke fordi at, det er ikke for at være nihilistisk, men evidensniveauet på meget af det vi laver er altså ikke ret højt, nødvendigvis. Men man ved det virker, og så skal man passe på det hele ikke går op i noget, der kan måles og vejes, mange ting man gør, så er det jo fordi man ligesom har på fornemmelsen at det her, det virker og det her går godt, og det er ikke det hele der kan skrives ned på artikelform. Der er meget erfaring i sådan noget, og hvad er det for noget, det med næse for den her sygdom, jeg tror der er mere af det end som så, der er meget lægekunst stadigvæk. Og det er der også i den daglige praksis. Så det er godt vi har noget at støtte det op omkring, men der er jo ikke noget af det her, det er jo ikke geniale tanker, vel, altså det får jeg ikke Nobelprisen for. Det er jo rent metodemæssigt kvalitet, det er et godt produkt, men det er jo rent faktisk bare et håndværk, man kunne ligeså godt være tømrer eller murer. Den geniale tanke får man når man går ude i skoven om søndagen, og man går helt alene eller sammen med sin hund, og der er ikke nogen at snakke med. Og så går man egentlig og tænker over nogle ting man har gået og puslet lidt med, og så får man lige pludselig en eller anden idé. Det er det, der gør forskellen på rigtig forskning og håndværk. Hvis man går tur i skoven, det kan også være badekarret, eller hvad ved jeg. Jeg tror mere det er den vej rundt det kommer, eller i dialog med nogen. Men håndværket er godt, så bliver du nødt til at bringe det videre, hvis du skal have det i New England Journal of Medicine, Gastroenterology eller Pain for den sags skyld. Der skal være orden i det, hele vejen fra begyndelsen og opad, det er så håndværket. Men det er ikke der den laterale tanke opstår.

Question: *Hvad så med sammenhængen imellem forskning, det som du et eller andet sted kalder håndværket, og så de artikler der kommer ud i sidste ende. Er der en sammenhæng imellem kvaliteten der, altså kan man godt skrive en artikel der er god, og her tænker jeg ikke så meget på rent sprogligt men..*

Response: metodologisk?

Question: *...ja, på baggrund af et knap så godt studie?*

Response: Narjh, men vi laver jo ikke dårlige studier her.

Question: *Så måske ikke så meget dig...*

Response: Det kan man jo ikke, det kan jo ikke betale sig at lave det, hvis det ikke er i orden. Men jeg vil da godt indrømme at nogle gange publicerer vi data... vi sidder i øjeblikket og afbilder nogle data, de er fem år gamle, vi har været igennem møllen fem gange. Det er fordi vi har en ny tilgang til dem, og så skal vi jo ligesom have det ud også, og vi arbejder meget med andre metoder. Men nej, man kan ikke skrive en god artikel på dårlige data. Det skal man helt lade være med. Det skal man ikke spille tid på, så hellere lade data ligge.

Question: *Men vil du mene der er nogen der gør det?*

Response: Ja da, uha uha, men det er jo også en gradbøjning; hvornår er data gode? Hvornår tror du på dem, og hvornår tror jeg på dem? Men jeg vil sige de fleste de ved godt det ikke kan betale sig at bruge kræfter på det, men problemet er at vi altid har 50% flere opgaver end vi kan overkomme, og derfor må man jo vægte de ting ordentligt. Men hvis vi så har en ph.d., og de skal igennem det her studie, det skal jo også skrives. Så bliver man jo nødt til det. Derfor kommer der mange ting hvor man godt kunne synes at det kunne være bedre at lægge dem et sted hen. Men det er også svært at vide på forhånd. Men langt de fleste de skal være gode data, det skal det. Altså tag nu for eksempel... vi laver eksperimentel forskning her, har vi et eller andet mål, eller et lægemiddel, hvis der er en spredning der er større end middelværdien allerede på forhånd, eller intervariabilitet over tid, hvis vi giver lægemidlet en af gangene og vi ved det varierer sådan her frem og tilbage, så kan vi jo ikke bruge det til noget. Det skulle gerne være luget ud på forhånd. Men det er så bare ikke altid vi ved det, for det kan være, at pludselig så viser det sig at der kan være nogle tekniske ting der gør, at det og det mål det ikke er særlig godt, det er ikke særlig reproducerbart over tid, det finder vi ud af bagefter, og så bliver vi også nødt til at rapportere data, for det er også forkert ikke at publicere data. Det skal man jo gøre. Heldigvis da.

Question: *Det er jo en del af protokol-godkendelsen?*

Response: Ja, men det bliver de jo så bare ikke altid. Det er jo meget godt at hvis man har nogle data, hvis de er halvdårlige, eller specielt hvis de er negative, så bliver de jo aldrig publiceret, vel? Andet end i journal of junk, og det kan man jo ikke leve af.

Question: *Det kan man jo så ikke. Så ville jeg egentlig spørge dig om du vidste hvor ofte de er citeret de her artikler, men jeg kan jo se de allesammen er fra i år, så det er jo nok ikke så mange endnu...*

Response: Det ved jeg ikke, heller ikke om du havde spurgt mig om en der var ti år gammel. Det tjekker jeg egentlig aldrig med de her artikler. Jeg ser slet ikke rigtig på det. Jeg får nogle artikler der er gode, og mine egne dem følger jeg egentlig ikke rigtig op. Det er nok også for dårligt, men jeg synes egentlig hvis vi har lavet et godt stykke arbejde, så må vi se, så må andre bedømme det. Den artikel jeg har, der er mest citeret,

det er en artikel der hedder McGill-pain-questionnaire på dansk, hvor vi egentlig lavede en dansk, det var ikke en oversættelse, det er en måde at kvantificere smerter på, hvor man lægger forskellige ord ind. Man har forskellige ord, som man fandt i forskellige bøger, som beskriver smerter med, det er faktisk utrolig spændende. Så tog man 100 forskellige ord man fandt i litteraturen og lagde dem ind i nogle forskellige kvalitative grupper, nogle hører til det man kalder den sensoriske dimension, f.eks. sådan nogle hovedpineord, det er hamrende, bankende, dunkende, og så lagde man dem ind og så fandt man ud af, når man lagde dem ind på den måde, nogen de var jo evaluative, nogle var miscellaneous og sådan noget, og nogen de var affektive, og de ord de viste sig så at uanset om man var akademiker eller man var fejedreng, så kunne man bruge sådan nogle ord til at beskrive en brækket arm eller hovedpine med. Det er sådan at den del af sproget ligger genuint i ens opfattelse, det kan være det ligger i ens opvækst, eller det er nedarvet, hvad ved jeg, det er utrolig interessant. Men hvad der også var interessant er at man kan gøre det på tværs af kulturer. Man kunne oversætte noget fra italiensk til dansk, bare man gjorde det på den rigtige måde, eller fra amerikansk til dansk, så man brugte de samme ord, og vi kunne bruge den universelt. Det lavede vi så, og det var meget spændende, og jeg kom ind som et tilfælde, fordi dem der lavede det kunne ikke få lavet det færdigt, jeg havde snakket om det, og så overtog jeg styringen, og så blev jeg førsteforfatter. Det kan man sige var egentlig lidt tilfældigt, fordi de havde lavet det meste af arbejdet, de fik det bare aldrig rigtig gjort. Men så da vi så er så mange smerteforskere i Danmark, der er rigtig mange smerteforskere i Danmark, der er næsten altid en tre-fire artikler i Pain fra Danmark, og det var jo noget de allesammen skulle bruge, og så begyndte de jo at citere det, når de skulle bruge den her skala, så citerer de mig hver eneste gang, og jeg synes ikke det var særlig genialt, for jeg havde ikke tænkt ret mange tanker, vel, jeg havde bare skrevet artiklen. Den er vildt citeret, det er den mest citerede artikel jeg har, jeg har skrevet så mange andre gode artikler, og de er ikke nær så citerede. Det er jo rent tilfældigt, jeg har ikke fortjent det overhovedet. Så derfor ved jeg ikke rigtig, med det der med at citere...

Question: *Det var sådan set også bare for at høre...*

Response: ...men det siger sig selv, at den bliver meget citeret den her (1), de er allerede begyndt at citere den, allerede nu. Den udkom for en uge siden, og den har været i new York Times og den har været alle mulige steder. Den her vil blive noget citeret (2), nu er det ikke så hyppig en patientgruppe, men det er patienter med en sygdom der har et stort problem, den vil blive noget citeret, det er jeg helt sikker på, for man vil bruge det her, men det har mange bivirkninger, så hvor meget effekt det har, det ved jeg ikke. Den her (3) ved jeg ikke, den vil blive meget lidt citeret, fordi det er meget basal forskning, og der er ikke så mange der arbejder med det her, men derfor er kvaliteten af artiklen god.

Question: *Nu havde du jo et editorial med til I'eren, har det en særlig betydning, at I også har fået et editorial udgivet i sammenhængen?*

Response: Det er et kvalitetsstempel, der er jo ikke andet end een editorial.

Question: *Det er inviteret, går jeg ud fra?*

Response: Det er nogen der er kommet bagefter, ham der, vi kender ham godt...

Question: *Nåh, det er ikke jer selv?*

Response: Nej nej, de beder en anden een at skrive om det her, om problemstillingen, og så citerer de vores. Vi har faktisk fået fire editorials her i efteråret på vores artikler. Som regel får vi en hvert andet år, vi har også fået en editorial på den der(2) og på en anden og tredje. Det er sådan et særligt kvalitetsstempel, at man har bidt så meget mærke i det, at man skriver en editorial på det.

Question: *Er det noget, der er unikt for det sundhedsfaglige, tænker jeg? Er det sådan rimelig stringent at editorials på den måde er en slags kvalitetsstempel for en anden artikel, eller er det bare nogle gange, at det er det?*

Response: Alle tidsskrifter har en editorial, et tidsskrift som Pain har mange editorials, men de har også altid 50 artikler i, og så har de fire-fem editorials eller flere. Editorials highlighter emnet, og det er en appetitvækker til at folk skal gå ind og læse artiklen. Men jeg tror i virkeligheden ikke det foregår på den måde i dag. For i gamle dage da sad vi jo med New England Journal of Medicine, og havde lige en time til at sidde og læse hvad der var kommet af spændende ting, det gør vi jo ikke længere, det kan godt være nogen gør, men det gør jeg ikke. Jeg søger, hvis jeg har et emne, og så ser hvad der er skrevet om f.eks. Barrett's esophagus, og så vælger dem ud der popper op, vi holder jo slet ikke de tidsskrifter længere, de er jo kun elektronisk, sådan er det blevet. Man sidder ikke længere og får appetit ved at læse editorials. Selvfølgelig er der nogle tidsskrifter jeg læser, f.eks. Pain, det får jeg også, og Ugeskrift for Læger får jeg hver anden gang, og jeg læser det kun når det er i hardcopy, jeg får det ikke gjort ellers.

Question: *Det sidste jeg har det er en lille brainstorm, hvor jeg vil nævne nogle begreber, og så vil jeg gerne have at du svarer med det første du tænker på, selvfølgelig indenfor konteksten forskningskvalitet. Du må selv om hvor meget du vil sige, om det er positivt eller negativt, og så videre. Det første begreb jeg har er peer review.*

Response: Jeg gør det hele tiden, så jeg burde jo have en mening om det... Det er nok den eneste mulighed der er for at få en rimelig uafhængig vurdering af en artikel, der er så det problem ved det, at det er en frivillighedsproces, der gør at, da du altid er bagud med alting, at du har lidt dårlig samvittighed over at du også bliver nødt til at gøre det, når andre gør det ved dig. Derfor sidder du og gør det, hvor man maksimalt sætter et par timer af en aften, hvor man i virkeligheden burde sætte en hel dag af til det, hvis man virkelig skulle gøre det ordentligt. Og derfor har jeg det lidt med det peer-reviewning, at det kan godt være at det er en, der er så mange af dem efterhånden,

at det kunne godt være man skulle finde en anden facon. Så er det også det, at meget forskel på peer reviews. Nogle journals har jo fem reviews, andre har een, og nogle, sådan en som Pain, det er en frygtelig journal, du sender det til to reviewers, eller, de sender den til to eller tre, og så får du dit svar tilbage, at det ser godt ud, og så svarer du på det de vil, og når de så sender den tilbage, så sender de den til to andre, det er kun Pain der gør sådan noget. Og ligemeget hvem du spørger, det er ligesom at trække i lotteriet, for afhængig af hvem du sender den til, så får du forskellige svar. Så får du pludselig to nye, der mener noget helt andet, så bliver den afvist, eller også så skal du svare igen. Det er fuldstændig sindssygt som den journal opfører sig, det kan man ikke være bekendt. Men peer review er som at trække i lotteriet. Du finder en flinke mand, og så får du en pæn besvarelse. Jeg synes egentlig jeg altid er meget positiv, jeg er glad for at lave forskning, jeg vil gerne hjælpe dem med at gøre det bedre. Min indstilling er positiv, men jeg kender kolleger, der afviser 98% af det princip at det skal fandme ikke være godt. Og derfor er der meget lotteri i det, og det er det, der er mig mest imod ved det. Og vi ved jo godt allesammen, at du kan sende den bedste artikel ind og så få den afvist, simpelthen bare fordi du er røget i det forkerte lotteri. Og det er jo ikke godt nok, jeg har bare ikke lige noget, der kan erstatte det. Så skal det være et hold af professionelle peer-reviewers på de enkelte journals, hvor folk får penge for at gøre det og sætter den nødvendige tid af til at gøre det. Og hvis man skal sidde en lørdag, så er det jo ikke gratis, og det er jo der man kan lave det. Så må det jo koste det, det gør. Og det gør man jo også nogle steder - jeg peer reviewer også for nogle af de store forskningsråd i Norge og Spanien og hvor jeg nu ellers sidder, og laver peer reviews af deres protokoller, det bliver jeg jo så lønnet for.

Question: *Men det er forskningsrådene?*

Response: Ja, men artiklerne gør aldrig, journals gør aldrig.

Question: *Hvad med sådan noget som...*

Response: Jeg afviste en i dag, for jeg kunne ikke holde det ud, jeg havde så meget at lave, så jeg kan godt gøre det når det bliver december. Nu er det begyndt at ske indenfor fjorten dage, det kan jo ikke lade sig gøre.

Question: *Hvad med sådan noget som sådan nogle boards hvor man uddeler fondsmidler, der har de også noget review, har de ikke?*

Response: Jo, jeg har lige fået en stor CD fuld af ting jeg skal læse, det er helt umuligt. Der er 30 ansøgninger, eller 50 ansøgninger man skal læse.

Question: *Er det også lønnet?*

Response: Nej nej, det er helt gratis. Vi får en lille smule, vi plejer at få en småkage til. Der er nogen af dem der er lønnet, og så er der frynsegoder ved nogen af dem, ved jeg.

Det gode ved at sidde i sådan nogle boards er at du kommer til at kende nogle andre. Altså nogle gange gør man det fordi det er sådan noget man skal, for at være professor, så gør man det lokalt, men skal du sidde i et board et andet sted, så kan det godt være lønnet, med fondsansøgninger. Men som regel er det mere frynsen, og så kan man sige man får som regel en del ud af det selv, for selvom det er din egen gruppe der sidder der, og du kommer udenfor døren, så ved man også godt, at en af dem der sidder i boardet har en god gruppe, så der falder nok mange penge af til dig, når du sidder der. Det kan godt betale sig, derfor er jeg altid med i boards. Jeg tror det er en god investering, også for os selv. Men man bliver jo også nødt til at gøre noget. Meget af det vi går og laver er jo reelt bare gratis arbejde - jeg kan jo bare sige nej til alt det reviewing og boards, men det er jo noget, der ligger implicit i faget, og hvis det forsvinder... det hele ligger jo i fritiden, og sådan er det bare. Ellers må man finde et andet job, tror jeg, for det kan man jo ikke være bekendt ellers, man kan jo ikke sidde og modtage penge fra boards, og så gider man ikke reviewe de andre. Specielt når man bliver lidt ældre - i starten fik man jo bare penge, men når man selv bliver inviteret med i boardet, så kan man ikke sige nej, for man fik jo selv penge dengang man var 25 år eller 30 år.

Question: *Det næste jeg har er layout af artikler?*

Response: Jeg synes det betyder meget, og jeg synes de er for lidt sexy nogle af de artikler. Tiden er løbet fra at have sådan nogle kedelige artikler. Der skal være nogle frække kurver, der skal være nogle tabeller og farver, nu har jeg printet det her ud i sort/hvid, det er NEJM, flotte tabeller. Jeg synes der skal være nogle illustrative figurer, og jeg synes det skal gøres meget simplere end det tit bliver gjort. Jeg vil så sige at artikel 3 her bærer præg af at være en tung, statistisk artikel, der er godt nok nogle fine print, men hvor er den kedelig. Hvorimod nogle artikler hvor der er en lidt fræk figur, det gør vi tit, hvor der er en forklarende figur. Men den der, den er simpelthen bare så fræk, den figur jeg fortalte dig her i artikel 1, hvor man ligesom slår dem allesammen sammen, det kan man jo ikke, det kan ikke lade sig gøre at lave sådan noget her, men den åd de alligevel. Den er fræk. Sådan noget skal det være. Jeg tror vi skal lidt frem i skoene der, for det gør det lidt sjovere at læse, videnskabelige artikler behøver ikke være kedelige, så kan man jo bare springe metodeafsnittet over, det er der jo ikke nogen der læser alligevel. Det er med at få nogle frække figurer frem. Så jeg synes det betyder noget med layout, helt klart.

Jeg sidder også i nogle editorial boards og diskuterer det her, hvor vi diskuterer hvordan tidsskrifterne skal se ud. Så det er ikke fremmed at diskutere det her. Jeg plejer også altid at være lidt mindre konservativ end de andre, det skal være lidt frækt. Du kan se hvad der sælger aviserne; de vil se sport, så skal der helst være nogen der går nedenunder og hjem med noget økonomi, og så skal der være noget sex, og det er det, der sælger. Der er ikke nogen der gider læse andet.

Question: *Det næste jeg har er citationer, det har du jo næsten svaret på.*

Response: Det har jeg nemlig, med index, h-index og sådan noget, ja. Men der er igen det der med citationer at når folk bliver kendte, så bliver de meget citeret fordi de er kendte, der er meget at man går efter dem der er kendte på bjerget i det her, og de bliver meget citeret. Men selvfølgelig betyder det da meget, og det er også derfor vi prøver at få det i de bedste journals, for de bliver citeret mest. Og det er jo klart at er man meget citeret bliver man også inviteret til de rigtige ting, og så får man også nemmere ved at få sin næste artikel antaget, og dem fra gruppen har, så det betyder noget.

Question: *Impact?*

Response: Jamen jeg tror det har meget betydning for os, vi går altid efter højeste impact. Vi laver meget basal forskning, og kan vi få det i American Journal of Physiology, som har en impact factor på 2, så er vi ovenud tilfredse. Ligeså tilfredse som vi er den anden dag hvor vi får en i NEJM. Men det kommer an på hvad type [artikel] man skriver, men det er da klart at impact det betyder noget. Vi er da mere stolte af dem der har høj impact factor end de andre. Men jeg synes man rider lidt for meget på det. I forhold til hvad det reelt er værd.

Question: *Vil du drage en parallel imellem tidsskrifters impact og så de artikler der er deri?*

Response: Jamen det er klart, men det kommer jo an på hvad det er for artikler. Hvis det er kliniske artikler, så får de et højt impact. Hvis det er basal videnskab, så har det lav impact, per definition. Så selve faktoren den vil jeg ikke, men jeg vil sammenligne indenfor et område, f.eks. smerteområdet, Pain har højest impact factor indenfor smerte-journals, efterfulgt af Journal of Pain, efterfulgt af European Journal of Pain, efterfulgt af nogle af de andre, så de har allesammen en rangorden, og det er da klart man går tit og ofte, hvis det skal være indenfor smerte... men vi gør jo så tit det vi vælger en anden journal, f.eks. Anaesthesiology, der også er meget smerte-relateret, den har højere impact, og den bliver nok læst af flere klinikere, så jeg tror ikke kun det er det, men det er klart at de grupperer sig afhængig af område. Skal vi lave en epidemiologisk artikel går vi efter de høje impacts fordi vi har sådan nogle gode databaser her i Danmark, det er meget nemmere at få en epidemiologisk artikel i NEJM. Vi prøvede også at få den der i, sådan en der vil ikke gå i, det skal være noget der fuldstændig radikalt ændrer vores måde at tænke på, det er også derfor det har så høj en impact factor. De tager jo kun dem... men du ville jo aldrig få en eller anden basal smerte-artikel i NEJM. Det kan ikke engang lade sig gøre. Vi kan være heldige at få den i Lancet på en rigtig god dag. Men det er meget usædvanligt, Nature måske, eller Cell. De tager basal forskning på højt niveau.

Question: *Så der er høj-impact-tidsskrifter der også tager grundforskning?*

Response: Ja, man kunne godt få en smerteartikel i Nature. Det kunne man godt, og Nature har en impact factor på 32 eller sådan noget, den er høj. Så det findes, men som

regel ligger det på det andet niveau. Og vi har prøvet at sende noget til Nature, det er meget svært, du har meget svært ved at komme igennem med noget som helst.

Question: *Tværfaglighed?*

Response: Det er lige præcis det vi gør mest i på det her center, af alle i verden, tror jeg. Vi er jo kendt i hele verden for vores tværfaglighed. Det er jo det vi har, vi har 8-9 faggrupper fordelt på de 25 mennesker vi har, så det er næsten kun to af hver. Så det er forskellen, vi prøver at tænke på den samme måde og uddanne hinanden så vi kan forstå hinandens fagområder; vi trækker ting fra ingeniørvidenskab, fra matematik og lægevidenskab, vi har to dyrelæger siddende, vi har cand.scint.'er siddende, og vi har to farmaceuter siddende, og vi har sygeplejersker siddende, vi har så mange faggrupper. Og lige præcis i vores tilfælde er det tværfagligheden der gør forskellen i forhold til de andre grupper. De andre grupper der er rundt omkring i verden indenfor det vi gør, de er alle sammen, langt de fleste, en eller to fagområder. Vi er så mange fagområder, og vi prøver at respektere hinandens fagområder. Man er ikke bedre her fordi man er læge eller ingeniør eller omvendt. Så det er ekstremt vigtigt, specielt i Mech-Sense.

Question: *Det sidste begreb jeg har, det er ny viden*

Response: Det er jo det vi genererer hele tiden. Man kan sige en masse om ny viden, det er jo det, det hele handler om. Jeg vil sige at ny viden og ny udvikling, det er ikke bare viden, det er også innovation, at man får nye redskaber. Men det er klart, ny viden er interessant hvis det er rigtig viden, der er meget ny viden der ikke er rigtig. Lad os nu tage den her [artikel] nr. (1), nu har den fået noget ny viden om Barrett's esophagus, det kunne være vi havde lavet en regnefejl, så havde det ikke været særlig godt, for nu holder vi op med at kontrollere alle dem her for kræft, og så får de kræft allesammen. Nu er det Danmark, det får de ikke. Men selvfølgelig skal man passe på med det, og der er meget der er populistisk. Og så er der det med al ny viden, at det går ligesom en EKG, plejer jeg at sige. Altså du har en lille p-tak, så går det lidt i bund, så går det helt op, så skal alle have det, så skal ingen have det, så til sidst så går det lidt på det jævne, så er der lidt der får det. Det er sådan jeg synes det går hver gang. Det går sådan for det meste af ny viden tror jeg, men det er godt nok. Det er jo fantastisk at arbejde i medicin, vi kan smide alle de lærebøger væk der er fem år, for der er kommet så meget ny viden de ikke kan bruges mere.

Question: *Det var sådan set det jeg havde, medmindre du har et eller andet?*

Response: Næh, ikke andet end hvad du skal bruge det til.

Question: *[Followed by explanation of PhD project.]*

A.2.6. Interview reference: M09

Date: Oct 18, 2011

Time: 14:00

Question: *Hvis du vil starte med at fortælle lidt om de artikler du har taget med, især med henblik på kvaliteten i dem, hvorfor har du vurderet dem til at være høj kvalitet?*

Response: Jamen jeg har været så ubeskedent at tage to artikler fra min egen afdeling, og det er jo fordi jeg kender dem godt, og jeg ved hvad de står for og omhandler. De er begge to fra *Annals of Internal Medicine* som er karakteriseret ved at være et godt blad, med en høj impact factor og også nogle interessante artikler, men også som har en høj standard for hvordan man afrapporterer videnskab. De har som regel været med i det når der bliver lavet nye guidelines for afrapportering af f.eks. observationelle studier, som der blev lavet her for nylig. Så har de været et af de blade, som har trykt de her guidelines, og som har rettet sig efter dem. Og faktisk er det sådan at den her artikel (2) med Lohse som førsteforfatter, den er faktisk brugt i deres blad som instruks for hvordan en god artikel den kan sættes op. Så det synes jeg jo er meget flot at opnå det, at blive brugt som eksempel på hvordan det kan gøres. Så artiklen er god, og det er jo fordi den for det første omhandler et meget vigtigt forskningsspørgsmål, nemlig om hvordan folk der bliver nydiagnosticeret med HIV, hvordan deres overlevelse er i forhold til baggrundsbeholdningen, og der har man jo nogle idéer om at det nok har ændret sig lidt fra før man fik HIV-behandling og så til senere perioder. Og der synes jeg den her artikel er god, fordi den beskæftiger sig med det her forskningsspørgsmål, og fordi den på en elegant måde udnytter de datakilder vi har i Danmark, der kan belyse det her. Altså, man kan følge, man kan finde en stor og komplet kohorte af nye HIV-patienter og så følge dem over lang tid, f.eks. ti år frem i tiden, i vores registre, så kan man udtrække en matchet gruppe af baggrundsbeholdningen, som man også kan observere over lang tid og sammenligne med de her HIV-patienter. Og der kan man oprette en sammenlignelighed på aldersprofiler og forskellige ting, så de her grupper ligner hinanden udover HIV-infektionen. Det synes jeg er godt og kreativt tænkt, og det er et vigtigt forskningsspørgsmål. Og så er artiklen meget velstruktureret og velformuleret, for det er egentlig et meget vanskeligt emne, statistisk set, hvordan man har gjort det, men det synes jeg artiklen formår at formidle meget flot. Og så har den nogle gode tabeller og især en meget god figur her sidst i artiklen, som senere er blevet brugt rigtig meget, når man snakker om HIV-patienters overlevelse. Så god figur, gode tabeller og så velskrevet og så følger den strukturen i STROBE (von Elm et al., 2007) kriterierne, som hedder *Strengthening the Reporting of Observational Studies [in Epidemiology]* som følger en bestemt måde at stille sin artikel op på, hvor man først ridser baggrunden på og så beskriver man sine inkluderede patienter meget nøje, så hvem der er ekskluderet og hvorfor, hvilket setting man arbejder i, hvilke eksponeringer og outcomes man kigger på, hvilke counfounder forstyrrende faktorer og hvordan man tager højde for dem, alt

det her er meget klart beskrevet.

Question: *Nu ser jeg lige at der er en lille notits her fra Editors. Er det standard at der er det i *Annals of Internal Medicine*?*

Response: Ja, det er der nemlig.

Question: *Er der mange andre tidsskrifter hvor man ser det?*

Response: Ikke mange. Men jeg vil nok sige en stigende del af tidsskrifter, hvor man har sådan en lille boks med "Why is this important"-agtigt. Der er en række tidsskrifter der har nogle hovedbullets opstillet i en boks med "What this study adds to the literature" og "Why is this important" og sådan noget. Og *Annals of Internal Medicine* vælger så at gøre det på en lidt anden måde med: hvad er konteksten, hvad bidrager det her studie med, hvad skal man lige passe på når man fortolker, altså cautions, hvor de så skriver at Danmark har en formiddabel HIV-behandling med fri og gratis adgang så man skal passe lidt på med at generalisere de her resultater til alle mulige andre steder. Og så nogle implikationer til sidst. Og det synes jeg er alletiders. For mange af os når jo aldrig videre end til lige at skimme abstractet og når der er sådan en boks så kan man lige fange hovedessensen, og det at de har cautions på en meget profileret plads siger også noget om at de er meget grundige. Man kan faktisk også se i selve abstractet der er der heller ikke ret mange blade, der ligefrem har et afsnit, der hedder limitations. Men det skal man udfylde, når man laver sit abstract. Så man kan ikke komme udenom meget eksplicit at skrive noget om de vigtigste limitations, det synes jeg er godt.

Question: *Vil du mene det er en ekstra kvalitet ved artikler i *Annals of Internal Medicine* at de har det her, sammenlignet med tidsskrifter hvor der ikke er det?*

Response: Ja, det synes jeg helt klart. Jeg kan rigtig godt lide deres måde at strukturere deres abstract fordi man, jeg vil ikke sige tvinger, men vejleder folk kraftigt i hvad man skal huske at stille op, for det giver nogle bedre artikler, simpelthen. Jeg bedømmer jo på forskellige tidsskrifter, og der er indimellem et stort rod og en stor pærevælling. Her tvinger man folk til at strukturere lidt. Så derfor kan jeg godt lide at læse artikler i det her blad.

Question: *Er der noget konkret ved dem, hvor du vil sige "det her, det er afgørende for god kvalitet."*

Response: Tænker du på artiklerne her?

Question: *Ja. Men man kunne også bare sige ved forskningsartikler generelt.*

Response: Ja, jeg synes det er vigtigt at det er et vigtigt forskningsspørgsmål, og det må meget gerne være en original idé, både med hvad man vil undersøge, men også måden

man undersøger det på. Metoderne må gerne være originale. På den måde at det ikke lige er noget der er gjort før. Det synes jeg kendetegner en god artikel. Det er også tit det de gode tidsskrifter leder meget efter. De vil gerne have et eller andet novel, noget der ikke er gjort før på den her måde.

Question: *Hvad så med de studier, nu kan jeg ikke lige huske... EPIC er vist eet af dem, hvor man har nogle meget store datasæt og så anskuer man en problematik fra X forskellige vinkler, og så udgiver man en artikel for hver X, eller for hver vinkel. Kommer vi så over i noget salami-forskning, eller er det også god kvalitet?*

Response: Jeg kender ikke lige det eksempel du nævner...

Question: *Det er f.eks. tilfælde hvor man er gået sammen i et multicenter studie, eller man har bare taget et stort datasæt, det kunne være fra Danmark, og har anskuet en problemstilling... lad mig finde på et eller andet... noget med kost, hvilken effekt kan det have på hjerte-kar-sygdomme f.eks., så kunne man starte med at se på noget med rødvin, og man kunne se på noget med fedtsyrer, man kunne se på mange forskellige ting. Hvis man så skriver een artikel per emne man behandler, altså en om rødvin, en om fedtsyrer.*

Response: Ligesom i det Europæiske kost-kræft-helbred studie, som eksempel, der gør man jo netop det, man etablerer et kæmpestort datasæt og så bruger man det til at belyse en masse forskellige problemstillinger. Og det synes jeg egentlig er på sin plads, at man etablerer de her meget store datasæt, som man putter en masse ressourcer i. At man så også virkelig får noget videnskab ud af det. Jeg kan godt høre hvor du vil hen, det er lidt den diskussion med salamisering af ting, versus at lave meget omfattende artikler med forskellige problemstillinger. Der er jo betydelige problemer ved at lave nogle artikler der omfatter en hel masse problemstillinger, en masse eksponeringer og en masse outcomes i samme artikel. For det første bliver det lidt noget rod, der er en fare for at man ikke rigtig kommer i dybden med de her associationer, fordi associationen imellem f.eks. rødvin og hjerte-kar-sygdomme, der er en hel masse litteratur specifikt på det område, og der er en hel masse stoffer i rødvin, der kan netop have betydning i forhold til vaskulære outcomes, så alene diskussionen af det, hvis man ser på hvor meget plads der er i et baggrundsafsnit i en almindelig artikel, eller et diskussionsafsnit også, det er en ret stor, omfattende litteraturgennemgang og diskussion det kræver. Hvis man skal gøre det for ret mange associationer i samme paper, så bliver det jo ret hurtigt uoverskueligt og man når ikke rigtig i dybden med noget af det. Noget andet er det metodemæssige, hvis det nu f.eks. er rygning som risikofaktor for hjerte-kar-sygdomme, så hvis man lavede artiklen specifikt på den problemstilling, så kunne man gå i dybden med rygning som eksponering, man kunne prøve på at identificere passiv rygning, tidligere og nuværende rygere, man kunne lave noget med rygemængde og man kunne faktisk gå i dybden med en masse ting omkring den eksponering, der hedder rygning. Hvis du forestiller dig at hvis man næste gang laver en artikel med alkohol og med kost som eksponering, så er det salamisering, for du har jo spurgt om alle de her ting, hvorfor laver du ikke een stor

artikel der præsenterer det hele? Så vil de igen metodemæssigt og analysemæssigt være fuldstændig vildtvoksende, og skulle gå i dybden med alle de her eksponeringer. Så de artikler du her foreslår; store, omfattende artikler, de vil tit blive sådan af en karakter; rygning: ja/nej, alkohol: ja/nej, ny eksponering: ja/nej, fordi man simpelthen ikke har plads til at gå i dybden med noget af det. Så det bliver nogle mere overfladiske artikler.

Question: *Så det bliver en balancegang imellem problemstilling og data og hvor meget der findes om det, hvordan man kan definere problemstillingen?*

Response: Ja, det synes jeg. Det bliver en balancegang. Og så er der det med massesignifikans også; hvis man undersøger mange problemstillinger, så er der jo en ret stor sandsynlighed for, at hvis man undersøger tyve sammenhænge at een eller to vil være signifikant sammenhængende bare på grund af tilfældigheder. Det er selvfølgelig ikke så meget anderledes, når man skriver tyve enkeltartikler på de tyve associationer, eller om man præsenterer dem i een artikel, så det synes jeg egentlig ikke er noget godt argument. Man må forholde sig kritisk til sine data, og hvis man undersøgte mange associationer i een artikel må man selvfølgelig fortælle læseren at nogle af de her ting kunne opstå ved tilfældighed. Det vil man så typisk ikke gøre, når man tager salami-metoden og undersøger mange associationer, fordi der er hver association ligesom taget fat på frisk, på ny. Men faren er sådan set den samme, og vi kommer stadigvæk ikke udenom at det giver noget bedre videnskab, hvis man går i dybden med nogle sammenhænge. Og så er der jo noget med kausalitetskriterier, hvor det ene ekstrem er data-mining, hvor man i et stort datasæt går ind og lader computeren undersøge flere hundrede sammenhænge og finder noget der bonger ud og så skriver på det. Det vil jo være noget der i mine øjne nærmer sig svindel. Den anden model er at man har læst nogle spændende hypoteser om at rødvin kunne give den ene eller den anden effekt, så tager man sit datasæt og går ind og undersøger specifikt den sammenhæng. Det er jo den måde man bør arbejde på, altså hypotesedrevet. Problemet for læseren er at det kan være svært at afgøre hvad baggrunden har været, når man læser et studie.

Question: *Ja, når man bare har artiklen?*

Response: Ja. Det kunne jo være det var noget, der var poppet op ved tilfældighed, og så lader forfatteren som om det er en hypotese de længe har gået og ville undersøge. Og litteraturen er så omfattende, at næsten uanset hvor skør en hypotese er, så kan man finde et eller andert, der kan underbygge det. Så det er med at bevare sin kritiske sans og få gentaget resultaterne med nogle andre datasæt i nogle andre settings. Et enkelt studie kan aldrig nogensinde stå alene.

Question: *Så reproducerbarhed er også en vigtig kvalitet?*

Response: Ja, det er en meget vigtig kvalitet. Jeg siger nogle gange, når de ph.d.-studerende kommer til mig og siger, "Åh nej, nu har de i Boston lavet vores studie, så er

alt fortabt, så kan vi lige så godt pakke sammen.”, så siger jeg, “Prøv nu og hør, for det meste er det faktisk sådan at et studie som det her øger opmærksomheden og interessen for at der laves et nyt studie, der kan be- eller afkræfte den her ting. Man behøver ikke nødvendigvis være den første, det er selvfølgelig skønt, men det er altså ikke sådan at når man har lavet eet studie, især et observationsstudie, så lukker det yderligere studier. Tværtimod.”

Question: *Hvad med sådan noget som studiedesign, har det noget med kvalitet at gøre? Nu tænker jeg også i forhold til evidensniveau.*

Response: Det er jo sådan noget jeg underviser i, og det har det jo selvfølgelig. Men noget af det, der er vigtigt at huske på, det er at det rigtige, eller det bedste studiedesign, hvad det er, det afhænger af forskningsspørgsmålet. Vi kender allesammen simple evidensgradueringer, A, B, C, D og så videre, og som regel er det de randomiserede trials eller meta-analyser af randomiserede trials, der står øverst. Men den tabel gælder sådan set kun for tilsigtet behandlingseffekt, altså studier der undersøger behandlinger. Og hvis man tager prognosestudier, som vi meget laver indenfor min afdeling, så er det faktisk et andet evidenshierarki, så er det faktisk store kohortestudier, der er den bedste evidens, og ikke randomiserede trials. Det er der flere grunde til; dels er der flere selekterede personer i trials, så generaliserbarheden er ofte ringe for prognosen på de folk der er med og i almindelighed, dels er studierne ikke store nok til at finde alle mulige outcomes og hvad kan man mere sige, ja, for det tredje er folk ofte ikke fulgt i lang nok tid til at man kan finde mere sjældne outcomes, ligesom cancer eller andre ting der kan opstå. Så der er altså flere grunde til at kohortestudier og observationsstudier de er bedre end trials. Til nogle forskningsspørgsmål.

Question: *Er det almindelig kendt blandt læger, at der er flere forskellige evidensskalaer? For jeg hører dem ikke nævne mere end en.*

Response: Nej, det er det nok desværre ikke, det er ikke mit indtryk i hvert fald. Så derfor tager jeg rundt og underviser i det. Jeg underviser i forskningsmetode og det er nok nødvendigt. Også i nationale guidelines i patientbehandling, der ser man som regel kun præsenteret den her evidensgraduering, der vedrører behandlingsstudier, men der er altså mange der.... indenfor klinisk epidemiologi siger man “patienten har mange spørgsmål til lægen, der ikke specifikt berører behandlingen. Hvorfor har det her ramt mig? Hvordan vil min prognose være?” - og det kan simpelthen ikke besvares ud fra et trial. Al etiologisk forskning, altså rødvin/hjerte-kar-sygdom kan ikke undersøges i et trial, det er simpelthen ikke feasible, det er ikke etisk, det er ikke muligt at lave et stort trial hvor man randomiserer folk til at de skal drikke fire glas rødvin om dagen, og nogle andre, der slet ikke må. For det første kan det ikke lade sig gøre, for man kan ikke holde folk randomiseret i lang tid på den måde, for det andet er der etiske problemer, og for det tredje er der ressource-problemer. Så der er masser af forskningsspørgsmål der ikke er egnet til et trial. Eller du kan tage, da man begyndte at observere de underlige

immundefekter hos homoseksuelle i USA først i 80'erne, der kan man jo ikke opsætte et trial med at være homoseksuel eller ej, og finde ud af hvad det her handler om. Der laver man selvfølgelig en case-serie som det første, hvor man beskriver det her syndrom, og som det næste step så laver man en case-kontrol undersøgelse, hvor man sammenligner folk der har fået det her nye syndrom og folk der ikke har, og så undersøger man deres risikofaktorer og deres profil. Som det tredje kan man måske lave et kohorte-studie af, lad os sige homoseksuelle og heteroseksuelle som man følger over tid. Der er sådan et naturligt hierarki hvis man vil finde ud af noget indenfor medicin. Så det her med straks at smide et trial på bordet, det er kun ganske bestemte forskningsspørgsmål det egner sig til, det synes jeg lige man skal huske.

Question: *Jamen jeg synes det er rigtig interessant, fordi jeg har spurgt mange af mine andre deltagere, om det er selve designet der er afgørende for kvaliteten, eller om det er sammenhængen imellem spørgsmålet og designet. Og det har du jo faktisk netop lige selv svaret på, og du svarer det modsatte af hvad de gør. Og det synes jeg faktisk er rigtig interessant, men det er måske også fordi du netop sidder her, og de andre sidder som klinikere langt hen ad vejen, og det er måske det, der gør forskellen. Det synes jeg er interessant i hvert fald. Hvis man nu skulle prøve at evaluere kvaliteten af forskning, objektivt gerne, så fremt det er muligt. Kan man evaluere den, og hvordan, hvis man kan?*

Response: Ja, det er et stort og svært spørgsmål... jeg ved ikke om du kender bogen "How to read a paper" af Trisha Greenhalgh (Greenhalgh, 2010)

Question: *Jo.*

Response: Den er faktisk rigtig udmærket. Der er nogle anvisninger til hvordan man kan læse artikler. Og noget af det første hun anbefaler er, om det er det rigtige design der er anvendt i forhold til forskningsspørgsmålet. Hvis det ikke er det, så kan du ligeså godt skrotte paperet, uanset om det har nogle interessante resultater. Så det skal man selvfølgelig finde ud af, når man evaluerer kvalitet. Man skal være klar over sit studiedesign, om det er passende, og så skal man være klar over, om det design så er anvendt rigtigt. Altså, er der mange nok patienter med, eksempelvis, til det man vil undersøge? Er grupperne velbeskrevet? Hvad med forskelle imellem grupperne? Det er som regel to grupper man vil sammenligne. Er de forskelle velbeskrevet? Er de håndteret statistisk? Har man kunne justere for de forskelle, hvis det nu er et observationsstudie? Har man kunne gøre det tilstrækkeligt? Har der været en klar analyseplan og har forfatterne sandsynliggjort at resultaterne er valide i forhold til det de har gjort med resultaterne? Der er ikke nogen nemme genveje. Så må man gennemgå en artikel for at finde ud af om den er god eller ej.

Question: *Men sådan nogle ting som du nævner her, er det ikke noget der et eller andet sted, i en idéel verden, burde blive fanget i et peer review?*

Response: Jo, det er det. Det er jeg enig med dig i. Nu er det jo ikke kun metodefolk og epidemiologer som jeg selv, der laver de peer reviews. Det er ofte kliniske kolleger, der ikke har en tung baggrund indenfor observationsstudier, der ikke ved hvad de skal kigge efter. Det kan godt være de gør, det er jo væsentligt at sige, at der er mange klinikere der er veluddannet, også indenfor forskningsmetode, men det er der altså også mange der ikke er, og som synes en eller anden sammenhæng er helt vildt interessant, og det har de også hørt om et andet sted, og så begynder man måske at nørde rundt i en masse mekanismer, og det er også fordi cellerne er sådan og sådan, o.s.v., men man glemmer sådan set det grundlæggende, nemlig om analyserne er lavet rigtigt. Mange blade er begyndt at spørge om reviewerne mener, der skal være en statistisk evaluering af de her analyser, og spørger reviewerne om de selv føler sig i stand til at reviewe om analyserne var gode nok, eller om de eventuelt skal have fat i en statistikers bedømmelse. Nogle blade vil også have en statistiker i deres reviewpanel og nogle metodefolk, så de kan sørge for, at der måske i hvert fald er een der er dygtig til de her ting. Det vil jo være smart, synes jeg i hvert fald, at have.

Question: *Hvad med... hvad tænker du om idéen om en professionalisering af peer-review, at man i stedet for frivillig basis som det er nu, ansætter nogle folk til det?*

Response: Det ved jeg ikke rigtig. Jeg har hørt det foreslået, ganske sporadisk, i en eller anden sammenhæng. Men jeg har ikke selv gjort mig nogle nærmere tanker om det var en god idé. Man kunne nok argumentere for, at fagligt ville det være en idé, hvis man kunne finde nogle metodemæssigt dygtige folk, og samtidig nogle folk, der er uafhængige af conflicts of interest i forhold til... for det er nok de to væsentligste problemer ved det nuværende system; dels at man har folk der ikke oplyser deres interessekonflikter, der synes at alt hvad der kommer fra den diskussion er noget hø, og som samtidig ikke har kompetencer til egentlig at vurdere tingene rigtigt. Så det kunne man selvfølgelig omgå på den måde. Men der er nok også nogle ulemper. Nu har jeg ikke tænkt så meget videre over det, men jeg synes jo det har en uddannelsesmæssig værdi også, at du reviewer som forsker. Man får øvet sig. Det fungerer jo typisk sådan, at man får fra sin professor eller fra sin vejleder, måske et nemt manuskript, som man skal prøve at bedømme, og så kommer man i dialog med sin vejleder om hvordan man egentlig bedømmer videnskab. Det synes jeg er en meget god idé. Sådan nogle ting ville jo falde bort, hvis det professionaliseres meget. Det giver også mulighed for en interaktion med kolleger. Online-tidsskrifterne er jo begyndt at have open-review i en høj grad, hvor man kan se hvem det er der bedømmer, og det kan der også godt komme noget godt ud af.

Question: *Det kunne måske være med til at øge kvaliteten af det?*

Response: Ja, jeg synes i og for sig det har nogle gode elementer. For man giver sig måske lidt mere umage, og man er knap så hurtig på aftrækkeren med at jorde noget på et eller andet løst grundlag og interessekonflikter bliver tydelige på den måde. Der

vil man nok sige nej hvis man reelt har en konflikt. Så ja, man kan jo nogle gange tage kontakt til en reviewer og få nogle gode råd, til hvordan man kan gøre noget bedre, så man får denne faglige interaktion, som jeg synes er god. Så kunne man måske som en mellemting være lidt mere omhyggelig med hvem man vælger som reviewer til sine ting. Noget hvor man har en vis kvalitetskontrol over det. Der virker det jo sådan nu, at de store blade er attraktive at være reviewere for, for det kan folk skrive på deres CV, og det ser godt ud man er bedømmer på *Lancet* f.eks., for så har man nok en idé om, hvis man er blevet spurgt om det, så er man nok ikke den allermest grønne, håbløse forsker et eller andet sted. Og sådan er det nok også i praksis. Så det bliver nok de større, lidt mere velfunderede forskere, der sidder og evaluerer hinandens ting ofte. Men igen - man skal jo også have nogle comming-ups, nogle folk der kan tage over. Det dur ikke, at det kun er grand old men der sidder og bedømmer hinandens ting. Så jeg ved ikke hvordan man skal sørge for at kvaliteten er god, det er slet ikke så nemt. Så skulle det være så attraktivt at folk de indsender et CV og en application for at blive bedømmere. Sådan tror jeg faktisk det er for nogle af de største blade. At man kan søge om at blive reviewer. Noget af det der sker lige nu er at det eksploderer helt vildt, med online-tidsskrifter. Når jeg ser på nogle af de mails jeg får fra kinesiske og indiske nye tidsskrifter - hvor skal de dog få bedømmere fra til alle de mange, mange blade og underlige artikler de får ind. Nu siger jeg underlige, det er jo lidt en fordom, men det er klart at kvaliteten vil være meget svingende i mange af de her nye online-tidsskrifter. Jeg tager mig selv til hovedet - skal vi nu også have endnu et blad om et eller andet medicinsk teknologi noget i forhold til de 25 der er i forvejen, fordi det her forlag også skal have en profil på det. Så siger det næsten sig selv, at man ikke får nogle ekstremt dygtige, kompetente reviewere, der står i kø for lige at være reviewer der.

Question: *Det sidste jeg har er en brainstorm, hvor jeg vil nævne nogle begreber og hvis du så bare fortæller mig hvad du tænker når du hører dem. Du må selv om hvad du vil sige, men selvfølgelig i konteksten forskningskvalitet. Det første det er ny viden.*

Response: Ja... jeg tænker ny viden, er måske nummer et kriterium for ethvert forsknings spørgsmål - er det her nyt? At der ikke er tredive studier der har undersøgt det samme, eller at man har nogle gode, konklusive studier i forvejen. På en eller anden måde skal det her være nyt. Og det er også det der præger de her tekstbøse, "What's new here?". Og det er egentlig meget god anledning til at overveje, "Er det egentlig nyt og relevant, det jeg laver? Eller laver jeg det bare fordi jeg kan lave det", sådan bør det ikke være.

Question: *Tværfaglighed?*

Response: Ja... i min branche er tværfaglighed når det er godt et samarbejde imellem metodefolk, der ved noget om designs, og klinikere der ved meget om sygdommene, og så statistikere der ved rigtig meget om analyser, der synes jeg tværfaglighed kan være rigtig godt. Det kan også være en almen mediciner der sidder i et andet setting, hvis

man har den del med i sin forskning. Sådan tænker jeg tværfaglighed derfra hvor jeg sidder.

Question: *Er det vigtigt?*

Response: Jeg synes måske det er lidt et buzz-word nogle gange, det er lidt fortærsket, at nu skal vi også være tværfaglige. Og det skal vi vel ikke for enhver pris - vi skal bare have nogle folk med, der ved noget om det vi laver. Vi behøver vel ikke nødvendigvis have en fysioterapeut med også, fordi vi kigger på kirurgiske patienter, når de folk der i forvejen er med faktisk også har en viden om træning eller kirurgiske komplikationer. Så man skal ikke være tværfaglig for at være tværfaglig.

Question: *Impact?*

Response: Ja. Impact er jo vigtig - altså forskning der rykker, forskning der flytter vores måde at tænke sygdom på, tænke på sygdomsmekanismer, tænke sygdomsbehandling, forskning der ændrer behandling, der har direkte implikationer for patienter. Det er forskning der har impact. Og så er impact selvfølgelig også impact factor i forhold til tidsskrifter, som er væsentlig når man har en forskerkarriere som jeg har. At publicere i tidsskrifter, hvis artikler citeres meget.

Question: *Det næste begreb jeg har er også citationer.*

Response: Citationer er jo vigtige, og tiltagende vigtige. Når man skriver ansøgninger i dag, så er det næsten et krav at man giver det totale antal citationer f.eks. de sidste X år på ens artikler, eller gennemsnitscitationer på sine artikler, h-index og den slags, som du jo selv arbejder med. Det er jo blevet tiltagende vigtigt, det kan jeg godt se.

Question: *Mener du at det afspejler kvaliteten?*

Response: Det synes jeg nok det gør. Især hvis man gør, som man gør nu, og som er en god ting, synes jeg, og prøver at bruge forskellige mål og kombinere nogle forskellige mål. Hvor jeg nok synes at det her med gennemsnitlig impact factor f.eks. på ens publikationer, det har været lidt sort/hvidt og forsimplet. Det er jo netop ikke noget der siger noget om hvor meget mine artikler så er blevet citeret, hvor stor impact min forskning har haft, hvorimod citationer til min egne artikler, der er der en klar sammenhæng imellem om det er nyt og andre finder det væsentligt. Så det er godt at kombinere forskellige mål, synes jeg.

Question: *Det sidste jeg har er layout af artikler.*

Response: Jamen man kan jo sige at der er flere planer. Der er jo layout af manuskript, som man sender ind, som betyder meget for præsentationen og om det lykkes at få noget accepteret. Og så er der selve layoutet som det ser ud her i artiklerne når de er

trykt. Jeg prøver at lære vores ph.d.-studerende, at det er vigtigt med layout når man sender noget ind. Det er vigtigt at lave nogle pæne tabeller og nogle flotte figurer. Ikke et eller andet hvor linierne de hopper frem og tilbage med skiftende mellemrum og så videre. Det giver indtryk af sjusk. Og det smitter negativt af på bedømmelsen. og så er der jo en stor tilfredsstillelse i at få et lækkert layout sådan som i det her blad, med lidt farve og lækre skrifttyper og tingene stillet pænt op. Det siger også en lille smule om tidsskriftets niveau måske. Men der har man jo nye muligheder nu, med online-tidsskrifter hvor det tilsyneladende ikke koster ret meget at lave et flot layout. Det synes jeg da er imponerende hvad selv små online tidsskrifter... og så får man en utrolig flot artikel tilbage som man kan distribuere rundt.

Question: *Det var sådan set det jeg havde... nej vent, lige en enkelt ting, ved du hvor ofte disse to artikler er blevet citeret?*

Response: Nej, det ved jeg faktisk ikke, jeg har ikke noget tal på det, men det kan jeg selvfølgelig gå ind og slå op.

Question: *Det behøver du ikke, det er sådan set bare fordi jeg gerne vil vide om du har været inde og slå dem op.*

Response: Joeh, jeg har været inde og slå den her (2) op, fordi jeg for nylig har skulle slå mine artikler op i Scopus, som du har hjulpet mig med engang. Og der har den nok været citeret... jeg kan ikke huske om det er 30 til 50 gange. Den har i hvert fald været meget citeret. Og det vil jeg egentlig også forvente, for det er et emne med miltfjernelse, som er meget vigtigt. Denne her (1) har jeg så ikke kigget på, for den har jeg ikke selv været med på, men jeg vil gætte på, den er citeret endnu flere gange. Og så spiller det selvfølgelig også en stor rolle med årstallet. Altså for os trods alt stadig lidt yngre forskere er det jo svært at få et højt h-index fordi publikationskurven jo gerne skulle accelerere lidt for sådan en som mig, så det vil sige rigtig mange af ens artikler er måske kun to-tre år gamle, så de kan jo ikke være blevet citeret flere hundrede gange, hvilket de måske har om fem år, de gode af dem.

Question: *Det var sådan set det, medmindre du har andet at tilføje?*

Response: Nej, det synes jeg ikke.

A.2.7. Interview reference: M10

Date: Oct 17, 2011

Time: 10:00

Question: *Hvis du vil fortælle mig noget om hvorfor du har taget de her artikler med?*

Response: Ja, jeg har taget to artikler med, og det er to artikler som jeg selv har læst og også brugt i mit eget forskningsarbejde og den første artikel (1) er en meget citeret artikel, hvor man er gået ind og har set på nogle væksthæmmede fostre, hvordan man håndterer når man opdager de er væksthæmmede inde i mors livmoder - om man venter til de vokser sig lidt større eller om man tager dem ud øjeblikkeligt. Den timing af forløsningen. Og det er lavet som et multicenter studie hvor det er 69 hospitaler i 13 lande, så der indgår 548 gravide kvinder. Så har man set på outcome alt afhængig af om de bliver forløst øjeblikkeligt eller om man venter lidt. Og den er publiceret i *Lancet*, og den har været meget afgørende fordi man fandt ud af det faktisk er bedre at lade dem blive lidt ældre, uanset at de ikke har det særlig godt, forstrene. Så den har ændret ved vores holdning til når man opdager et barn er væksthæmmed og måske ikke har det helt godt inde i livmoderen. Så var man tidligere nok mere aggressiv for at få det barn og få det i en kuvøse og give det noget mad i kuvøsen, hvor man nu siger at barnet i hvert fald ikke skal ud for tidligt. Så den har ændret vores praksis betydeligt, og den er blevet citeret mange gange. Så den synes jeg er rigtig god, og den er som sagt også fra et meget velestimeret tidsskrift, og det er også nogle rigtig gode afdelinger, der er med.

Den anden artikel (2) er et review om det samme emne, for det er det jeg interesserer mig meget for. Også med fostre med væksthæning. Hvordan deres udvikling, deres neuroudvikling, er, når man ser det i forhold til hvordan de havde det inde i livmoderen, nogle af de parametre man kan måle, bl.a. ultralydsblodgennemstrømning og sådan nogle ting. Og der er der sådan en meget dygtig læge, der hedder Baschat, der kommer fra USA, der har lavet et meget omfattende review hvor han, synes jeg, har alle de vigtige ting med. Så det har også været en artikel som jeg har læst meget og som jeg også har brugt som udgangspunkt, hvor jeg har læst referencerne for at komme omkring de vigtigste ting jeg synes der er indenfor det her felt. Så jeg synes den er meget grundig og velskrevet og rigtig god. Den er også fra et udmærket tidsskrift, som hedder *Ultrasound in Obstetrics & Gynecology*, som ikke er nær så godt som det første men som er mere et lille specifikt tidsskrift som bringer nogle rigtig gode artikler indenfor det her område. Så de er gode de to.

Question: *Vil du prøve at fortælle mig lidt mere specifikt, altså, er der nogle bestemte ting ved de artikler som du vil sige er afgørende for at det er god kvalitet?*

Response: Generelt vil jeg sige, også når jeg søger, så kigger jeg på hvilket tidsskrift de er publiceret i. Og så ser jeg lidt på om de er nye og så ser jeg også på forfatterne, for jeg har læst så meget indenfor det her felt her, så jeg har nogen jeg synes skriver godt og som jeg godt kan lide at læse noget af. Og så kan man jo sige at en ting er et review, der kigger jeg rigtig meget på forfatteren, men hvis det er sådan et studie her så kigger jeg også meget på designet, om det er en randomiseret, kontrolleret undersøgelse sådan som den her er. Og hvor mange der er med. Og når jeg er på conference, så bliver der også citeret fra de her store, internationale studier - så skriver jeg det op og læser det når jeg kommer hjem. Så det er også noget med hvad jeg har hørt om.

Question: *Det med konferencerne - er det f.eks. når nogen har lavet en præsentation?*

Response: Ja, f.eks., på mange konferencer, hvor der f.eks. er nogen der har en lille case-ting, med fem f.eks., så kan de godt henvise til at "vi ved jo godt at sådan er det", og så bruger de et studie som det her, for at bekræfte noget af det vi allerede ved, og så skriver jeg nogle gange de referencer ned som jeg kan fornemme man er enig om at det er sandt, eller hvad man skal sige.

Question: *Vil du mene at det er en anden måde at finde studier på end f.eks. at lave en informationsøgning i PubMed, eller hvis du bare finder referencer i en anden artikel?*

Response: Jeg gør egentlig alle tre ting. Jeg noterer mig hvis jeg hører citeret en artikel som jeg synes lyder spændende og som jeg ikke kendte, så kan jeg godt finde på at skrive den ned, og så ser jeg også på referencerne til sådan et godt review her, og så går jeg også ind og søger, for der kommer jo også løbende noget nyt. For dem der bliver citeret til konferencer de er jo ofte adskillige år gamle, for det tager jo lidt tid inden man kommer igang med at citere, det tager jo noget tid inden der bliver skrevet nye artikler hvor det bliver citeret, så jeg synes egentlig det er nogle forskellige måder at søge viden på. Jeg synes, altså, jeg er jo lidt autodidakt udi det her forskning, men nu er jeg jo startet på ph.d. og har også været på kurser hvor jeg har lært lidt mere om litteratursøgning, men jeg tror nok ikke jeg er så dygtig til at søge i PubMed. Jeg får tit en meget stor sværm af artikler, hvor jeg så skal bruge noget tid på at sortere det ned, og der kan godt forsvinde noget i den proces når jeg sidder og søger, så der kan jeg godt lide at have sådan et intensivt review her, så jeg sikrer mig der ikke er nogle ting som jeg ikke har fået med. Og jeg tror det handler om min evne til at bruge søgemedierne som jeg ikke helt... jeg kan godt lide at komme gøre sådan noget for at sikre mig at jeg kommer hele vejen rundt.

Question: *Du nævner sådan noget som hvilket design det er som en ting der kan være afgørende for kvaliteten...*

Response: ...ja...

Question: *...er det sådan noget som evidens du tænker på der?*

Response: Ja, altså, der er jo nogle gange man bare samler nogle cases ind og beskriver dem og så kan man jo lave sådan et rigtigt studie hvor man laver et randomiseret design og sammenligner de to grupper og det får jo en langt større styrke kan man jo sige.

Question: *Du siger, at det ene tidsskrift er bedre end det andet, hvordan tænker du det?*

Response: Ja, det er jo sådan at f.eks. *Lancet* og *Ultrasound in Obstetrics & Gynecology*, det er jo sjovt at sammenligne de to, for *Lancet* er jo et alment tidsskrift, der dækker alle læger, og de optager artikler som de mener har almen interesse, så det skal være sådan at

det er noget der kommer til at betyde noget væsentligt, ikke nogle små nørdede detaljer, men noget større, væsentligt. Hvorimod *Ultrasound in Obstetrics & Gynecology* de har en lille niche af folk der virkelig ved meget om et lille felt, og vil gerne have en artikel med om en lille teknisk ændring som kan optimere en eller anden metode. Hvis det lige præcis er det lille felt man sidder med, så kan *Ultrasound in Obstetrics & Gynecology* jo være godt, men det der virkelig ændrer verden, også indenfor vores felt, det finder man jo nok i *Lancet*. Og der sidder jo et panel og vælger ud hvad de vil publicere, og de vælger jo noget som de tror på har en eller anden form for banebrydende element, eller i hvert fald håber på det. Og det er jo også når man selv sidder og skriver artikler, så er der jo det med impact factor at man kan læse det der tal, som handler om hvor mange gange man citerer artiklerne, og der er impact factor på *Lancet* langt højere end på *Ultrasound in Obstetrics & Gynecology*. Men derfor kan man jo godt have en artikel der egner sig langt bedre til... altså, det kommer an på hvem man vil nå. Men jeg vil da sige, at hvis jeg sidder og søger og ser den er fra *Lancet* så kigger jeg da lige lidt ekstra på den.

Question: *Så hvis du har et søgeresultat med f.eks. mange hundrede artikler, hvis der så står Lancet på en af dem, så giver du den mere opmærksom, end hvis der står noget andet?*

Response: Ja, for det man også kan sige, som regel, så vil de have læst den grundigt igennem, og der vil ikke være nogen fejl i designet eller statistikken, altså de er jo tjekket grundigt igennem og er meget valide.

Question: *Validitet er det også en form for kvalitet?*

Response: Ja, det synes jeg. Men der er jo også en gang imellem hvor de trækker den tilbage igen, fordi de er hurtigt fremme, så man skal nok altid læse dem kritisk, uanset hvor de kommer fra.

Question: *Så det er ikke nogen garanti?*

Response: Det er det ikke, men sandsynligheden er jo meget større.

Question: *Hvis vi prøver at brede det lidt mere ud, og i stedet for at se på de her to artikler, hvis jeg bare siger forskningskvalitet, hvad tænker du så på?*

Response: Jamen så tror jeg også jeg tænker meget på mit eget arbejde med forskning. Jeg har i hvert fald tænkt over, at nu har jeg arbejdet med forskning siden 2005, og jeg synes selv jeg har fået mere kvalitet i min forskning, både på den måde jeg strukturerer, f.eks. mine designs. Jeg har de her skanninger hvor jeg giver ilt til moren og ser hvordan fosteret reagerer på det, og i starten havde jeg ikke nogen helt entydig protokol for hvor lang tid de skulle have ilt, og jeg har fået det meget mere standardiseret nu, og jeg indsamler også mine data og opbevarer dem meget mere struktureret.

Question: *Så det er selve forskningsprocessen...*

Response: Ja, det tænker jeg også når du siger kvalitet. At man starter med at have styr på det, så man ikke sidder og bliver i tvivl om hvad det her er...

Question: *Gør det en forskel i forhold til slutproduktet, altså artiklen?*

Response: Ja, den er meget nemmere at skrive, når man har nogle mere entydige - fordi når man har noget lidt mere rodet noget, så kan man blive lidt i tvivl om hvor man skal sætte cut-off og det hele bliver sådan lidt mere... der skal helst ikke være for meget overladt til at man selv kan tage et valg, det skal helst være det samme hele tiden. Det gør det i hvert fald meget nemmere, og sådan er mine data nu. Det gør det meget nemmere at gruppere det og arbejde med det. Og mere sandt, fordi man ikke kan sidde og skubbe lidt rundt med det, fordi det enten er i den kasse eller i den kasse. Sådan er mine data nu, det synes jeg er rigtig skønt.

Kvalitet er også noget med hvordan man behandler data. Hvordan man får slutproduktet ud og også får det skrevet, formidlet.

Question: *Du nævnte det her med design før og vi har også snakket en lille smule om det. Men jeg er lidt nysgerrig, er det selve designet, studietypen, er det dét der er afgørende for om det er en kvalitet, eller er det sammenhængen imellem problemstillingen og designet?*

Response: Du mener hvis der er en eller anden problematik jeg gerne vil vide noget om? Så kan jeg finde et case-study som nøjagtig har beskrevet det jeg vil vide noget om, eller jeg kan finde et randomiseret studie som ligner lidt, hvad er det så jeg synes der er bedst?

Question: *Ja, for eksempel.*

Response: Ja, det er ikke så længe siden, hvis jeg nogle gange i min kliniske hverdag har en patient med et eller andet, hvor jeg så tænker at det var noget jeg gerne vil vide lidt mere om, og så finder jeg noget der ligner lidt, som er stillet ordentligt op og så har jeg fem cases som er præcis det jeg har. Og der kan de cases jo godt have stor interesse fordi de er nøjagtig det man sidder med, på den anden side så har det jo ikke den styrke, så man kan ikke sætte procenttal på, f.eks., men derfor kan det jo stadig være meget anvendeligt i situationen, og der er jo også mange tilstande hvor man ikke kan lave et rigtig dobbeltblindet, randomiseret, kontrolleret studie, hvor det er noget meget sjældent, og det er noget, der bliver samlet sammen over tid. Men jeg ved jo godt at hvis man skal se på de evidensgrader så er det jo selvfølgelig det randomiserede studie som er det højeste, men det er bare ikke altid muligt at få det. Og så tager jeg også cases med, hvis jeg ikke kan finde noget videre.

Question: *Men det lyder på dig, og ret mig endelig hvis jeg tager fejl, som om du lidt sætter - måske ikke et lighedstegn - men i hvert fald er der en meget tydelig sammenhæng imellem*

evidensniveau og så kvaliteten af studiet?

Response: Man kan jo også godt lave et studie hvor indsamlingen ikke er i orden og hvor grupperne ikke er i orden, så kan det godt se ud som et randomiseret studie, men det kan jeg jo sådan set ikke vide når jeg læser det. Så der kan jo godt være nogle fejlkilder alligevel - i et studie som kalder sig et rigtigt design. Der kan f.eks. også være et stort frafald som de ikke forklarer, eller det kan være en gruppe som ikke repræsenterer den jeg sidder med, som ikke er en generel gruppe, men en lille, selekteret gruppe, så man får fejl i det alligevel. Så der kan stadig være mange fejlkilder. Men man er jo godt begyndt, kan man sige, hvis man har tænkt sådan et design godt.

Question: *Hvis man nu skulle evaluere kvaliteten af forskning objektivt, så vidt det er muligt, kan man det? Og hvis man kan, hvordan?*

Response: Det er nok rigtig svært. Som du også siger, er der nogle designs som vi traditionelt anser for at være god kvalitet, men de kan jo stadig komme i miskredit hvis man ikke udfører dem rigtigt. Jeg tænker, at hvis den er gået igennem til *Lancet*, så har den ikke de store fejl på den måde, så havde de nok sorteret den fra, så hvis det er et velestimeret tidsskrift, så har de nok hjulpet med at finde fejl. Men det ved jeg jo også godt, sådan er det jo ikke altid, så er der artikler der bliver trukket tilbage fordi der er fejl i dem.

Question: *Men det er trods alt en form for kvalitetstjek?*

Response: Ja, det tænker jeg, at der sidder nogle rutinerede mennesker i de her review-udvalg og i hvert fald kigger dem igennem. Men jeg ved ikke, det er lidt svært, ligesom æbler og pærer, at sige hvad der er bedst - og netop også hvad er det man vil undersøge? Hvad er det man vil vide noget om når man går ind og kigger?

Question: *Vil du mene der er en direkte sammenhæng imellem artiklerne og den forskning der ligger bag ved? Altså er det muligt at skrive en god artikel på baggrund af et dårligt studie, eller følges det ligesom ad? Så hvis du har en god artikel, så er det fordi der også ligger et godt studie bag?*

Response: Der er ingen tvivl om at det jo er en særskilt evne at skrive en god artikel. Jeg synes man bliver bedre til det, jo flere man har skrevet. Så jeg tror ikke de to ting hænger helt entydigt sammen, men der er ingen tvivl om at det er nemmere at skrive en god artikel hvis du har noget godt materiale. Men jeg tror godt man kan skrive en artikel, hvor man har noget godt materiale men som ikke kommer ordentligt frem, f.eks. Og jeg tror også godt man kan skrive en artikel hvor man ligesom dygtigt kan skrive udenom nogle af svaghedspunkterne og så håde det ikke bliver opdaget. Men det er klart, du kan jo ikke ændre på de data du har, men det er meget fortolkning, diskussion og perspektiv.

Question: *Men så lyder det på mig som om det er noget skriveteknisk noget, at hvis man er dygtig til at skrive, så kan man skrive sig udenom nogle problemstillinger - men er det så stadig god kvalitet?*

Response: Altså hvis du har nogle dårlige data, men du skriver det flot?

Question: *Ja...*

Response: Nej, man kan jo sige at kvaliteten af forskningen den... det er vigtigt at kunne formidle det jo, det er også en del af kvaliteten, men hvis man skal tale om den vigtigste kvalitet så vil jeg da mene det er de data der ligger der, hvordan de er. Men man kan jo ikke komme udenom, og sådan er det jo også på konferencer, der er nogen der kan komme ud over rampen, og andre kan bare ikke, det er da også en meget væsentlig del af forskningen, at formidle det, både på skrift og på konferencer. Jeg tror godt du kan have nogle rimelig gode og interessante data der aldrig kommer igennem hvis du ikke får formidlet det ordenligt og rigtigt. Så jeg tror ikke man kan sige det ingen væsentlig betydning har, det tror jeg faktisk det har. Det kan jeg jo også se når man skal publicere artikler - det er vigtigt det er skrevet ordentligt og på engelsk og formuleret ordentlig. Jeg tror det giver et indtryk af at dataene måske er lidt lemfældige hvis det er skrevet på et dårligt sprog, jeg tror på en måde man kobler det lidt som reviewer.

Question: *Så nævnte du lige før din kliniske hverdag - det kan selvfølgelig være svært at skille ad, det her, men vil du mene der er forskel på hvad du vurderer der er vigtigt, og som kvalitet, i en artikel, når du læser den som kliniker og når du læser den som forsker?*

Response: Ja, jeg bruger meget litteratursøgning og læse artikler i min kliniske hverdag qua min forskning, tror jeg. Det er meget sjældent jeg går ind og finder et opslagsværk fra 1997 og finder sygdommen, så går jeg ind i databasen og finder cases, så på den måde så bruger jeg det meget. Og det jeg så kigger efter det er jo en gruppe patienter som ligner min, hvad var outcome og hvordan håndterede man dem, mere sådan lidt opskriftagtigt; hvad skal jeg være opmærksom på her, hvad kan der gå galt og hvordan skal jeg behandle de her patienter. Hvorimod når det er i forbindelse med mit forskningsprojekt så tror jeg nok jeg er lidt mere... så går jeg også meget ind for at forstå de bagvedliggende mekanismer... det er ikke sikkert det er så vigtigt når jeg står i klinikken, så skal jeg bare vide "hvad kan der ske her og hvordan håndterer jeg det", der er det ikke sikkert jeg behøver vide noget om at det er fordi der går en eller anden process igang. Der tror jeg, jeg går meget mere i dybden for at forstå mekanismen i mit forskningsprojekt, end jeg gør i min kliniske hverdag. Så der er nok forskel på hvad jeg læser, og hvilke afsnit jeg læser grundigst.

Question: *Kan du se forskel på dig selv og de læger der ikke har en forskningsbaggrund i forhold til hvad de læser? Nu nævnte du f.eks. det med opslagsværker.*

Response: Ja, det tror jeg i langt højere grad at dem der ikke forsker eller søger i

databaser, det kan man jo godt gøre selvom man ikke forsker, men der er en tendens til at dem der ikke anvender forskning til hverdag, de bruger mere de traditionelle opslagsværker, som de måske læste efter dengang de læste på studiet, altså medicinsk kompendium og de her store opslagsværker. Hvis jeg nu gør det, så går jeg altid i referencerne, det tror jeg ikke er sikkert man gør, hvis man ikke har den her forskningsbaggrund. Så tror jeg man tager den holdning der står i den bog, og den kan ofte være lidt gammel, kan man sige.

Question: *Ved du hvor ofte de artikler du har taget med er citeret?*

Response: Det ved jeg ikke... meget... den her (1) er meget. Jeg tror hvis man er til konference så bliver den nævnt ti gange.

Question: *Men du har ikke været inde i Web of Science og se efter?*

Response: Nej, jeg ved godt man kan det, men det vil være flere hundrede gange, det er jeg slet ikke i tvivl om. Men den her review artikel (2), den tror jeg ikke er citeret særlig meget, den er fra 2011 så den er meget ny. Og igen, det er også mest originalartikler der bliver citeret.

Question: *Godt, jeg skulle bare høre om du tilfældigvis havde været inde og se.*

Response: Nej, det ved jeg ikke.

Question: *Det sidste jeg har det er en brainstorm, hvor jeg vil nævne nogle begreber for dig, og så skal du bare svare det første du tænker på. Selvfølgelig i konteksten forskningskvalitet, og du bestemmer selv hvor meget du vil sige. Det første begreb er peer review.*

Response: Ja, så skal jeg sige et eller andet til det?

Question: *Ja, er det vigtigt eller ej, er det godt, er det skidt, er det afgørende for kvalitet?*

Response: Det betyder jo... det er det her board af reviewers der sidder i et tidsskrift og sorterer tidsskrifterne. Det vil jeg sige er godt - selvfølgelig kan de jo også have en holdning der gør det meget svært at publicere noget, især hvis det er meget kontroversielt, der er jo nogen der bliver ofre for noget politisk, så det kan jo også have noget negativt. Men jeg mener at man er nødt til at have en form for gatekeeper-funktion. Jeg ved godt der kommer flere og flere netbaserede portaler, hvor der ikke er det. Men umiddelbart synes jeg det er godt.

Question: *Men det lyder lidt på dig som om det måske kan holde noget tilbage som måske er for nyt?*

Response: Ja, det tror jeg også det kan, det har der jo også været eksempler på i løbet af historien. Hvor der er nogen der har forsøgt at sende noget ind til *Lancet*, f.eks., der

var så banebrydende at de tænkte “det kan bare ikke passe”, så bliver det publiceret i et lille tidsskrift i stedet for, og så ti år efter kommer der en kommentar om, at det nok var en fejl - der var blandt andet en dansk kvinde der var med til, inden for reagensglas-behandling, hvor man tog æg ud, ultralydsvejledt, og befrugtede dem og lagde dem op igen, og den teknik med at tage dem ud ultralydsvejledt, den er hun faktisk den første der bruger, og den forsøgte hun at bringe i *Lancet*, men det tog de ikke. fordi jeg tror de mente det aldrig blev til noget. Så fik hun det så publiceret i et mindre tidsskrift, og efterhølgende har [Lancet] så skrevet en notits om at det var nok en fejl, der tog de fejl. Så der er nogle gange hvor tingene er så nye at de... eller, man er måske i en anden trend eller en anden bølge end den der lige hersker på det tidspunkt.

Question: *Sådan noget som layout af artikler?*

Response: Det har jeg ikke så meget holdning til. Altså, du tænker på de forskellige tidsskrifter, om der er nogen der gør det bedre end andre?

Question: *Ja, for eksempel.*

Response: Det eneste er at jeg hader at læse de artikler hvor de kommer på nettet inden de bliver skrevet som PDF, så har man bare en lang tekst og så figurerne til sidst, det synes jeg er lidt træls. Men generelt er det ikke sådan at jeg tænker at der er et tidsskrift der er bedre end andre. Jeg læser meget *Ultrasound in Obstetrics & Gynecology* så det er jeg vant til.

Question: *Citationer?*

Response: Det er også noget jeg kigger lidt på, hvis det er en artikel som er meget citeret, det kan man jo også se inde i PubMed når man søger. Så tænker jeg den er mere interessant. Og jeg ønsker mig jo også at mine egne artikler bliver citeret. Det er jo noget man går meget op i.

Question: *Er der en sammenhæng imellem citationer og kvalitet?*

Response: Man må sige en ny artikel er jo ikke citeret, f.eks., så det skal man jo i hvert fald huske - at der går nogle år inden den bliver meget citeret. Men generelt må man nok sige at det er der. Blandt andet GRIT-studiet.

Question: *Impact?*

Response: Det kigger jeg også på, både når jeg selv skal publicere, for at udvælge et tidsskrift jeg synes kunne være interessant. Jeg tænker at der i hvert fald er en større sandsynlighed for at det er en god kvalitet hvis der er en høj impact.

Question: *Tverrfaglighed?*

Response: Altså forskellige... f.eks. så arbejder jeg med MR-skanninger af fostre, og der har jeg mulighed for at publicere i obstetriske tidsskrifter eller i radiologiske. Og der kan man sige, der er mange artikler jeg ikke kan forstå at læse i et radiologisk tidsskrift, fordi de er meget tekniske. Så jeg arbejder vel med den tværfaglighed i mit felt. Det synes jeg er en udfordring. Det skal man jo tænke over når man skriver. Jeg har altid skrevet til obstetriske tidsskrifter, fordi det ligesom er det jeg kan. Men nu overvejer vi at skrive en mere teknisk artikel, hvor jeg så ikke skal være førsteforfatter, men for at få den der tekniske del med, som er interessant for de radiologer som skal udføre de her undersøgelser. For der er jo bare tværfaglighed og forskellige nicher af folk man skal kommunikere med, og det tror jeg man skal være bevidst om.

Question: *Tilfører det noget?*

Response: Til min forskning?

Question: *Ja, altså, tilfører det noget til forskning at man er tværfaglig? Gør det den bedre, eller er det ligegyldigt?*

Response: Jeg synes i min egen forskningsgruppe, der har jeg haft en biolog inde, en fysiker, en radiograf og en radiolog, og jeg synes det er utrolig positivt, fordi man har meget forskellig viden, og det er ikke altid man kan overskue perspektivet af den viden man sidder med. Bl.a. den metode som jeg har bygget hele min ph.d. op omkring, som er sådan en metode hvor man kan måle iltmætning, som er brugt indenfor hjerneforskning. Der tror jeg at dem der sad med hjerneforskning havde ingen idéer om hvilke perspektiver jeg kunne se i det, ud fra mine fostre. Det er meget kreativt synes jeg, at sidde med alle de her vidensting og så få det blandet på en eller anden måde. Og man kan godt føle man er kørt fast i et eller andet.... der var et tidspunkt hvor jeg havde mine dyreforsøg, så havde vi simpelthen behov for at have en golden standard som vi kunne måle MR-signalet op imod. Jeg vidste ikke hvordan jeg skulle komme til at måle iltindholdet i de her dyrefostre, og så var der en biolog der sagde at han havde en iltmåler han havde brugt til nogle andre forsøg. Så lånte vi den, og den vidste jeg slet ikke fandtes, og den kunne komme ind i MR-skanneren og ødelagde ikke vores signal, og den har aldrig været brugt til det før, men der opstod ligesom en mulighed der. Og den havde jeg aldrig fået, hvis jeg selv havde siddet og tænkt, eller havde spurgt dem jeg kender om det. Så jeg synes det er meget konstruktivt med flere, meget forskellige personer i en gruppe. Det har jeg i hvert fald oplevet med min gruppe, og det har været meget godt.

Question: *Det sidste begreb jeg har er ny viden.*

Response: Jamen ny viden er for mig dét det drejer sig om - at man skal finde ud af noget nyt. Det er da det, som driver een, at man enten ser det man plejer at se på en ny måde og bliver klogere, eller også at man ser nogle helt nye sammenhænge som man

ikke vidste, som er med til at udfordre de dogmer og sandheder vi tror på i dag. Så jeg synes det er væsentligt at forskning medfører noget nyt.

Question: *Det var det jeg havde, med mindre du har noget som du synes du mangler at fortælle?*

Response: Nej, det har jeg ikke.

Question: *Så vil jeg sige mange tak.*

A.2.8. Interview reference: M13

Date: Oct 12, 2011

Time: 09:30

Question: *Jeg har bedt dig tage nogle artikler med, og jeg kan se du har taget dem med som PubMed ID, kan det passe?*

Response: Ja, det har jeg.

Question: *Hvis du bare kan sætte et nummer, 1-3, på dem også.... godt, har du lyst til at fortælle mig hvorfor du har taget dem med?*

Response: Jo, det har jeg, lad os tage nr. (1), den handler om at der er nogle områder i hjernen man via PET-scanning har fundet ud af har med depressioner at gøre. Og ud fra den viden har man prøvet at sætte nogle dyb-hjerne-stimulations-elektroder ned i et korsvejdspunkt for depression. Og ved at behandle nogle patienter som har medicinsk refraktær depression, dvs. al konventionel medicinsk behandling og elektrochok, det fungerer ikke, så har man prøvet at behandle dem med dyb hjernestimulation og så i den her svært syge patientgruppe har man så formået at behandle med success fire ud af seks mulige. Det jeg synes der er rigtig godt ved det her studie er at rent videnskabeligt er det virkelig velopbygget, altså, det er ikke sådan at... der er mange studier hvor man får følelsen af, at det er et tilfælde, at man har fundet et eller andet, men her kan man virkelig fornemme at der har været nogle tanker omkring hvordan det hænger sammen, man har lavet nogle PET-scanninger, hvor man virkelig kan se hvad for nogle områder, der har med depression at gøre, og ud fra det har man så lavet en hypotese om at hvis man nu går ind og tager det her overaktive område - det kan man se på PET-scanningerne - og så hæmmer det med dyb hjernestimulation, så kan patienterne have en gavnlig effekt af det. Og så har man så gjort det i praksis og set at det rent faktisk virkede. Så rationalet for studiet er hele vejen op igennem virkelig velovervejet, og man fornemmer virkelig at investigatorene har gjort alt hvad de kunne for at finde huller i deres studie, altså, de har lavet utrolig mange tests undervejs for at finde ud af hvad der er denne behandlings styrker og potentielle svagheder. Man har ikke den følelse som man nogle gange har,

at ting bliver skubbet lidt ind under gulvtæppet, personligt bliver jeg utrolig, hvad skal man kalde det... ikke kritisk, mistænksom hvis man er meget hurtig til at lave en meget favorabel konklusion på relativt spinkelt grundlag. Jeg kan godt lide at de virkelig har vendt og drejet hver sten samtidig med at de også har turdet at gå ind i noget helt helt nyt. For det er meget nyt det de har lavet der, altså det er ikke sådan salami copy-paste forskning, så det studie kan jeg rigtig godt lide.

Question: *Nu siger du investigatorerne, er det forfatterne selv, eller er det en ekstern gruppe?*

Response: Nej, det er forfatterne.

Question: *Godt*

Response: Så det synes jeg er et rigtig godt studie, man har virkelig nået noget nyt, bragt den funktionelle neurokirurgi et markant skridt videre, men man har også samtidig ikke været for cowboy-agtig og bare kastet sig ud i det.

Question: *Man har tænkt sig om?*

Response: Ja man har tænkt sig om, og taget små skridt frem, men alligevel har man turdet tage skridtene. Så det er rigtig godt.

Studie nummer to plejer jeg at kalde brainbow-studiet, det er også et meget innovativt studie, hvor vi er gået over fra den kliniske til den 100% basale forskning her. Man har været i stand til ud fra det farvesystem der hedder rød-grøn-blå at tage nogle transgene mus og så få alle neuroner i hjernen til at udtrykke en kombination af rød, grøn og blå, og det kan f.eks. være rød-rød-rød eller rød-rød-grøn, og så ud fra den blanding af protein som de her celler så begynder at producere, så producerer de et fluorescerende protein som har nogle forskellige farver når det bliver udsat for lys ved en given bølgelængde. Så man er sådan set i stand til at, musens hjerne lyser simpelthen op i et sandt farveorgie, det er også derfor man kalder dem brainbow-mouse, men det er jo ikke kun fordi det ser smart ud, og det ser flot ud, men det er ligeså meget fordi man så er i stand til ud fra fotometrisk at måle hvad er farven for en neuron og følge det her neuron rundt i hjernen og følge dens aktioner fordi det her neuron det har en special farve, et farve-ID. Og det er også noget innovativt og nyt og det er simpelthen genialt, det de har lavet, og det har åbnet et helt nyt forskningsområde. Nogle gange sker der nogle metodemæssige udviklinger, så åbner der sig helt nye veje, og det her har været et studie der har gjort - jeg tror det er tilbage fra 2004 - og hvis man ser hvad de her brainbow-mus har været brugt til efterfølgende, det er ikke småting. Det har givet mulighed for at svare på nogle spørgsmål man ikke kunne svare på før.

Question: *Så de har et eller andet sted grundlagt en ny metode?*

Response: Ja, det er en helt ny metode de har lavet, som virkelig har rykket. For ellers er der nogle gange når nye metoder kommer til - tiden er jo den største dommer - så

troede man at det var noget helt nyt og det skulle åbne nye veje, og så viser det sig at, nej, det funkede ikke rigtig når det kom til stykket - de forventede applikationer de faldt lidt til jorden. Men det her er en helt ny metode som egentlig har flyttet noget.

Det sidste studie, nr. (3), det er en gut der hedder Koutcherov som førsteforfatter, og det er meget basal neuroanatomi. Det er ikke særlig sexet, hvis man skal se på det på den måde, det er ikke et studie der laver de helt store overskrifter, men det svarer på en meget overbevisende på spørgsmålet, hvordan den menneskelige paraventrikulære kerne oppe i hjernen er bygget op. Og det har været et kæmpe arbejde de har lavet, man er slet ikke i tvivl når man har læst den her yderst udførlige artikel fra ende til anden, at det er vitterligt sådan det hænger sammen. Altså sådan som de beskriver den anatomiske opbygning af den her kerne i mennesket, man er ikke i tvivl om at ja, det må være sådan den er bygget op, den her kerne, eller det her kerneområde i hjernen.

Question: *Fordi det er så detaljeret?*

Response: Det er en detaljeret beskrivelse og der er blevet brugt så mange forskellige metoder, det er ikke fordi jeg vil sige det er elegant, det er bare meget veldokumenteret. Og det jeg også godt kan lide ved den er at den type artikler har en tendens til at blive lidt ulidelige at læse på en måde, men de mestrer også sproget og formår at gøre den behagelig at læse.

Question: *Nu nævner du lige at det faktisk er flere metoder de bruger. Er det sådan at du vil sige det er metoder, der supplerer hinanden eller er det...*

Response: Det er nogle histokemiske metoder, dvs. de bruger både noget der hedder enzymhistokemi, dvs. der er nogle enzymer som nogle af cellerne udtrykker, nogle andre gør ikke, og så laver de nogle farvninger med de her enzymer og så har de også lavet nogle immunohistokemiske farvninger, dvs. man tilsætter nogle antistoffer som binder sig til nogle specifikke antigener i cellerne. Og så ved at tage en lang, man tager simpelthen hjernen af en hel del patiente man har fået, og så skærer man dem op i ultratynde snit i det her kerneområde, og så har man simpelthen haft 10 forskellige farvemethoder, og så har man simpelthen farvet alle snittene med en af hver farvemethode. Og så ligge snittene ovenpå hinanden, fotografere dem og så lade en computer analysere dem og tælle cellerne osv. så er man simpelthen i stand til at kortlægge hvordan de her forskellige kerneområder er bygget op. F.eks. den del der ligger op ad den laterale ventrikel, den udtrykker den og den og den immunohistokemiske markør, og den udtrykker ikke den og den markør, men den der ligger ved siden af, den udtrykker så den og den, men ikke den og den. Men de har de to tilfælles, og det gør at vi kan adskille dem fordi de er forskellige på den her måde. Det har ellers været et stort problem i det her kerneområde før, at hvis man har brugt for få farvemethoder, så har man slået nogle ting sammen som ikke skulle have været slået sammen.

Question: *Så det de egentlig har gjort det er at de har brugt én metode til at undersøge en*

lille del, en metode til at undersøge en anden lille del, og så kombineret det?

Response: Ja, kombineret det hele og ligesom set på fællesmængden af det, nu har jeg selv gået og arbejdet en del med sådan noget her, og det er et kæmpe arbejde, det tager utrolig lang tid og man ved man får ikke *Nature*- og *Science*-artikler ud af det her. Men alligevel så fornemmer man at forfatterne har gået op i det her med liv og sjæl og de svarer også på et rigtig godt spørgsmål. Og indenfor det her område da er det også en artikel, der får en del citateringer.

Question: *Indenfor det her specifikke område?*

Response: Ja, man kan ikke snakke om det område uden man lige skal nævne den her artikel. Så den får en del citeringer.

Question: *Okay. Så har vi snakket lidt om artiklerne, og det synes jeg har været interessant at høre. Men jeg vil også gerne høre hvad du tænker mere generelt når jeg bare siger ordet forskningskvalitet.*

Response: Når jeg tænker forskningskvalitet, så tænker jeg det skal være noget der bringer videnskaben videre, altså ikke kvantespring, men bare små skridt, altså svare på nogle decideret nye spørgsmål. Og så tænker jeg det skal være velunderbygget. Altså ens konklusioner skal være baseret på et ordentligt fundament. Fordi det jeg nogle gange kan være lidt skeptisk overfor og lidt nedladende kalder for salami-forskning, altså forskning for forskningens skyld, på en eller anden måde, af karriere-mæssige årsager skal man lave noget forskning nu, ellers kan man ikke komme videre i systemet, og hvordan slipper man nemmest om ved det her? Og nogle gange synes jeg folk de... altså forskning er jo også noget med at komme ud over... komme ud med sine resultater. Jeg synes nogle gange der er nogle resultater der bliver solgt som at være lidt mere sexede end de i virkeligheden er, altså man forsøger at gøre det til mere end det nødvendigvis er, hvorimod det havde været helt fint at sige at "vores resultater kan bære det og det og det", så synes jeg nogle gange det kan være lidt belastende at læse et fint studie, men så forsøger man at få det til at virke som om "vi har fundet kuren mod kræft", nej, I har bare formået at komme et skridt videre og få en dybere forståelse af hvad det er, den perspektivering der nogle gange kan komme synes jeg kan være lidt for meget. Så forskningskvalitet vil jeg sige er meget med at der er et ordentligt fundament, og man har tænkt man forsøger at bringe forskningen videre i en relevant retning. F.eks. synes jeg ikke der er så meget kvalitet i noget af det forskning jeg har set rundt omkring, at man udvikler en metode, og det kommer der nogle ganske fortrinlige arbejder ud af, man så laver man ligesom noget salami-forskning, hvor man bytter een komponent ud med en anden, og så gentager man hele møllen. Og så bytter vi en anden komponent ud og gentager hele møllen. Altså at du har 30 ph.d.-studerende og du tager alle deres ph.d.'er og lægger op, så finder du ud af at den eneste forskel fra den til den er sammensætningen af en overfladebehandling af en eller anden protese, og så er det fuldstændig identiske

artikler, hvor det eneste der ændrer sig er overfladebehandlingen af nogle proteser, f.eks., og så hvilke tal der er nede i statistik-området, og så en konklusion, men ellers er det mere eller mindre identiske artikler og ph.d.'er der bliver skudt af. Det synes jeg måske ikke der er så meget kvalitet i, så synes jeg det går over og bliver meget forskning for forskningens skyld.

Question: *Så tænker jeg lidt på, du nævner at du synes forskningen skal bringe noget nyt til, eller den skal bringe videnskaben videre. Hvad så med forskning der validerer anden forskning?*

Response: Det bringer også videre, det er med til at cementere, det er også rigtig god [forskning]. Al forskning skal kunne overleve tidens tand, på en eller anden måde, jeg synes det er helt - det er også nødvendigt, man stoler jo mere på en metode eller en konklusion som er blevet efterprøvet masser af gange i stedet for at det kun kan du' i det her ene laboratorie, så bliver man også lidt skeptisk. Så det er det, validering på den måde det er nødvendigt.

Question: *Hvad så hvis man skal prøve at evaluere det? Kan man evaluere forskningskvalitet objektivt, og hvis man kan det, hvordan?*

Response: Jamen det er jo meget svært at evaluere noget forskning objektivt, for problemet det er jo tit at dem du sætter til at vurdere det, de har jo også en eller anden forhånds- altså en erfaring af en eller anden art, så det som måske i virkeligheden er et fuldstændig genialt resultat indenfor et eller andet område, men hvis du så sætter en der ikke ved virkelig meget om det område til at evaluere det, så kan det være de slet ikke kan se det. Så det er meget svært at få nogen til at vurdere det på den måde. Men man kan sige i bagklogskabens klare lys synes jeg måske det er let nok at evaluere på en eller anden måde, for hvor mange gider citere de her artikler? Altså er det nogle artikler der bliver trukket frem i ny og næ eller ofte igen og igen, fordi ja, der var andre der også syntes det her var smart, og det var med til at danne et fundament for at komme videre, så i bagklogskabens klare lys synes jeg nok lidt man kan, til en vis grad. Er det en artikel der kan opnå nogle citationer, eller kan den ikke? Det tror jeg er vigtigt. For det med at vurdere forskning ud fra hvilket tidsskrift det kommer i, altså impact factor for et tidsskrift, det har utrolig meget med held at gøre, altså jeg har da set nogle artikler som er i et ikke særlig højt rangeret tidsskrift, men så har det jo flere hundrede citationer, men så er der andre artikler der kommer i meget, meget fine tidsskrifter, og der er ikke een der citerer dem efterfølgende, altså, de kom, de forsvandt. Så jeg synes det med at bruge impact factor det tror jeg ikke rigtig på, men citationsindex det er nok det der er bedst til at vurdere det. Det er måske den mest objektive måde at gøre det på, jeg ved det ikke, men det er i hvert fald en mulighed.

Question: *Hvad så med sådan noget som, og det er jeg faktisk lidt spændt på at høre specielt fra dig, fordi du er den første kirurg jeg snakker med - sådan noget som evidens, altså*

evidensniveauer på forskellige studier, er det også noget der på samme måde er interessant for jer, eller er det mere ovre i det medicinske?

Response: Nej, jeg synes det er meget vigtigt fordi, især indenfor de kirurgiske, eller, det er indenfor alle de medicinske og kirurgiske specialer tror jeg, at man kan have en eller anden metode og så kan man af den ene eller anden årsag have næsten et religiøst forhold til at det her, det virker bare. Og så hver eneste gang det går galt, så finder man en eller anden forklaring på det, og så andre gange, eller, hvis man prøver en ny metode, og det går galt de første gange man prøver det, så synes man det er en dårlig metode. Jeg tror det er vigtigt at have noget evidens på det man gør, for tit så er det jo ikke et spørgsmål om hvad der er sandt, det er jo ikke sort og hvidt, det er en grå mudret masse i virkeligheden. Skal vi vælge det her indgreb eller det her indgreb - det ene er måske en anelse bedre end det andet, statistikt, men det er ikke noget den enkelte kirurg eller enkelte afdeling er i stand til at vurdere - tit er man nødt til at have et multicenter-studie til at vurdere om det er den ene eller den anden vej. Så evidensniveau det er vigtigt, altså man stoler da mere på et studie med rigtig, rigtig mange inkluderede patienter, og randomiseret og så videre, det har man da evigheder mere tiltro til end et der er baseret på fem case stories, for nogle gange må man også bare konstatere at en blændende kirurg f.eks. er i stand til at få en suboptimal metode til at fremstå som meget god. Så evidensniveau det er vigtigt.

Question: *Hvad er det så ved, altså, hvis vi nu skal prøve at sætte - ikke lighedstegn - imellem evidens og kvalitet, men der er en eller anden forbindelse, går jeg ud fra...*

Response: Ja

Question: *Og hvad er det så ved evidensniveauet der er afgørende? Er det det specifikke design, eller er det sammenhængen imellem designet og problemstillingen der er afgørende for om det er god kvalitet?*

Response: Den skal jeg lige have en gang til.

Question: *Feks. så snakker man jo ofte om at randomiserede, kontrollerede forsøg de har et meget højt evidensniveau...*

Response: ...især hvis de er dobbeltblindede...

Question: *...især hvis de er dobbeltblindede, ja, og er det så dét, at det er et randomiseret, dobbeltblindet forsøg, der er en god kvalitet, eller er det dét, at der er en sammenhæng imellem det design man har brugt, og den problemstilling man ønsker undersøgt? Det at man vælger en case-serie, kunne det godt i nogle tilfælde være en bedre kvalitet, end at man forsøgte at klemme noget andet ind over?*

Response: Jeg tror at den golden standard man har, der er ingen tvivl om, at hvis

man kunne det i alle tilfælde, randomiserede, dobbeltblindede studier, hvor man har en problemstilling og så laver man et design og trækker lod ved patienterne når de kommer ind ad døren og så kører man på, fordi det der tit er problemet når man gør det bagudrettet, det er jo om der ikke har været en selektionsbias undervejs.

Question: *Så det jeg hører dig sige det er, at det egentlig er selve designet der egentlig er det der er afgørende?*

Response: Ja

Question: *Godt. Et lille spørgsmål; hvis man nu skal se på sammenhængen imellem forskningsartikler og så den forskning der ligger bag, er det så en direkte sammenhæng, eller kan man godt skrive en god artikel på baggrund af et halvdårligt studie?*

Response: Det er der masser af eksempler på, der er nogen der er i stand til at skrive rigtig lækkert og flydende, og på nærmest magisk vis er i stand til at snige sig uden om alle studiets svagheder, og på en eller anden måde så man ikke reflekterer over det, for de er nævnt, men bliver alligevel bortforklaret på sådan en eller anden plausibel måde, men hvis man lige siger hov og går tilbage een gang til og ser på metoden, så er det lidt ahhhh... så der er ingen tvivl om at god forskning, lige så vel som der er lavet et rigtig godt setup, og studie og god metode, så er der også meget i evnen til at skrive lækkert.

Question: *Men den artikel som du så nævner, der er velskrevet, men hvor de sniger sig lidt rundt, er det så også god kvalitet?*

Response: Nej, det er dårlig kvalitet.

Question: *Så kvaliteten af artiklen hænger sammen med kvaliteten af studiet?*

Response: Ja, det må det gøre endegyldigt. Men derfor skal der ikke være nogen tvivl om at et rigtig godt studie kan komme til at virke lidt halvfesent hvis det er rigtig dårligt skrevet. Det er der ikke nogen tvivl om. Så lige så vel som det er nødvendigt at lave et godt studie så er det også nødvendigt at være i stand til at formidle sine resultater på en fornuftig vis. F.eks. et af mine yndlingstidsskrifter, det er Journal of Comparative Neurology, det læser jeg hver gang det kommer ud, de siger bl.a. også at metoden er essentiel, så efter introduktionen kommer materiale- og metodeafsnittet, og de bruger faktisk meget plads på det, hvorimod der er mange andre tidsskrifter, der skrotter materiale og metode bagerst i artiklen og de må ikke bruge ret meget plads på det, så hvis du sætter dig ned og læser deres materiale- og metodeafsnit, så er du ikke i stand til at eftergøre deres forsøg, fordi det er så meget på floskelplan.

Question: *Kan du tilfældigvis nævne et enkelt tidsskrift hvor det er sådan?*

Response: Nej, men jeg har nogle artikler derhjemme, hvor jeg kan konstatere, fordi

jeg har forsøgt at eftergøre det, og det kan man ikke, for der mangler simpelthen for mange punkter i opskriften.

Question: *Hvis du husker det må du meget gerne sende en reference til mig. Jeg har faktisk ikke selv set det, men jeg synes det er meget interessant.*

Response: Altså når jeg læser en artikel så læser jeg abstractet til at starte med, for lige at se om jeg gider begrave mig i det her videre, og så læser jeg bagefter materiale og metodeafsnittet for simpelthen at se hvordan det er bygget op, og om jeg tør stole på det. For hvis man også har læst materiale og metodeafsnittet inden man læser resultatafsnittet, så er det meget lettere at lure om de sælger en vare de har dokumentation for eller ej.

Question: *Jamen jeg er helt enig, og jeg undrer mig også over man kan finde på at smide metodeafsnittet bagi.*

Response: Det er faktisk mange, jeg tror det er Nature der også er begyndt på det, eller også er det Science, og jeg synes jo selv det er kikset.

Question: *Okay. Så har jeg kun to korte ting tilbage, og det ene har du selv været lidt inde på. Har du en idé om hvor ofte de er citerede de tre artikler du har med?*

Response: Mayberg-artiklen....det er tocifret, og hvis du slår den op og finder ud af det er trecifret, så bliver jeg ikke chokeret. Der er ikke nogen indenfor det her område med dyb hjernestimulation for depression, der er det hjørnestensartiklen, den skal simpelthen nævnes i alt, mere eller mindre fra nu af til dommedag, for det var der det startede. Jeg tror det er ligefør du kan sætte et lighedstegn imellem alle artikler der omhandler DBS for depression, og så antal gange den er citeret, og så vil citationerne overstige antal artikler, fordi det er psykokirurgi og efter historien med det hvide snit er det noget der ... næst efter behandlingen for OCD, med DBS, er det et af de første helt store områder der er blevet kigget på, så det har også skabt en masse kontrovers, også her i landet. Der er flere artikler i Dagens Medicin, Ugeskriftet og også på nettet og jeg tror også i Politiken eller Jyllandsposten der også har haft det fremme. Så den har rigtig mange citationer.

Brainbow (2) tror jeg ikke har helt ligeså mange fordi de her mus ikke bare er nogen du kan få fingrene i. De er rigtig svære at få fat i, men den har også en hel del citationer, men det er ikke i nærheden af Mayberg artiklen. Men det er fordi det her forskningsområde, hvor Brainbow-musen ligesom har startet det hele op, det er jo ikke et kæmpe forskningsområde med multiple centre rundt omkring i verden - jeg tror det er et eller to centre der arbejder med den her mus, men til gengæld det der bliver spytet ud, de har den med hver gang, og de artikler hvor man har brugt de her mus som eksperimentelt dyr, de ender alle de gode steder, og de svarer vitterligt på noget. Så jeg tror ikke det er fordi den har så mange citationer.

Den sidste (3) er heller ikke fordi den har så mange, den har en 10-20 tror jeg, det er også godt synes jeg, fordi det man kan sige - den ene, første artikel, det er meget bredt, så den har en masse citationer, hvorimod de to andre er meget smalt, det er vitterlig nørdet, men det er rigtig godt. Og det kan man se ved at de områder der har med det at gøre har dem med hver gang.

Question: *Til sidst har jeg en brainstorm, hvor jeg vil nævne nogle begreber, og så vil jeg gerne at du siger det første du tænker på, selvfølgelig i konteksten forskningskvalitet - positivt, negativt, kort eller langt, det er helt op til dig. Det første begreb er ny viden*

Response: Positivt. Skal jeg bare sige positivt eller negativt?

Question: *Nej, du må gerne uddybe.*

Response: Ny viden det er positivt, det er med til at bringe os længere frem - ingen ny viden så stagnerer vi bare og kommer ikke videre. Men det skal være ny viden baseret på et solidt fundament og ikke bare luftkasteller.

Question: *Så ny viden er afgørende for forskningskvalitet.*

Response: Ja. Man er nødt til hele tiden at presse videre og ikke blive for selvfed.

Question: *Og det var jo faktisk noget af det første du sagde, da du fortalte om de her artikler, det var jo netop ny viden. Det næste er tværfaglighed?*

Response: Jeg tror sådan som verden er skruet sammen nu og forskning er skruet, det er at næsten al god forskning er typisk tværfagligt. Fordi den tid, der hvor der virkelig sker noget nyt, det er brydningsfladerne mellem flere forskellige specialer, f.eks. Mayberg-artiklen (1), hvor du har psykiatere, psykologer, neurologer, neurokirurger og neuroradiologer, mange specialer, der er gået sammen om det her, og der er ingen tvivl om at havde der manglet bare eet led i den her palette, så havde det ikke duet, det er derfor det også er så godt et studie, så tværfaglighed det er ved at være essentielt for den meste nye viden. F.eks. Brainbow musen er ikke tværfagligt, det er genetikere og biokemikere der har siddet og lavet det her, men det er også meget smalt. Det er genialt hvad de har lavet, men det er ikke tværfagligt. Men langt hen ad vejen, og især indenfor klinisk forskning, der tror jeg at fremtiden ligger i det tværfaglige.

Question: *Impact?*

Response: Altså du tænker på impact factor?

Question: *For eksempel.*

Response: Jaeh... det er svært at vurdere, for der er ikke noget tvivl om at ved mine egne artikler der er sigtet efter så høj impact factor som overhovedet mulig, som det

overhovedet kunne bære, og jeg synes det er fedt når man får det i et tidsskrift med en god impact factor indenfor ens område, og man bliver da træt af det hvis det ender i noget med knap så høj impact factor, men når det virkelig kommer til stykket så synes jeg nok det især er det med hvor mange gange det bliver refereret, det er nok det der på dommens dag virkelig tæller.

Question: *Men når det så er fedt at det kommer i et tidsskrift med en høj impact factor, er det så fordi det har en høj impact factor eller er det fordi det er et godt tidsskrift?*

Response: Det hænger sammen. Fordi når nu Nørre Snede Håndbold Klubs videnskabelige tidsskrift det... det er bare ikke cutting edge, og de tidsskrifter med høj impact, de artikler der kommer der, enten er det forholdsvis nyt, fundamentalt nyt, det bringer forskningen videre, eller også er det noget der tager en eksisterende hypotese, der står lidt på gyngende grund, og enten efterviser at det var sådan, eller i nogle tilfælde at nej, det var det ikke. Så jeg synes mange af de artikler der kommer i de store tidsskrifter, enten svarer de endegyldigt på om tingene hænger sådan eller sådan sammen, eller også er de i stand til at bringe det videre. Der er utrolig lidt som slipper igennem nåleøjet som ikke enten bringer forskningen fundamentalt videre med noget nyt eller afprøver noget, og slår det fast med syvtommers søm, så fast man nu kan slå det. Men jeg synes det er utrolig sjældent i de store tidsskrifter at der er noget der er sådan lidt nåeh, jaeh, joh, men svarer de så på en af de store hypoteser? Nej, ikke rigtig, den type artikler ser man meget sjældent, medmindre man så ser i forfatterlisten, og ser det er en af de helt store hotshots som er gode venner med nogen af editors, for de er de også, de artikler, hvor man så kan undre sig over hvordan den artikel kom med. Men det er ikke så tit, og jeg tror der bliver mindre og mindre af den slags, fordi tidsskrifterne også slås så meget med hinanden og beskytter deres impact factor. Det er jo også fordi forskningsmængden i dag er eksploderet i forhold til hvad den var for tyve år siden. Der er aldrig nogensinde blevet pumpet så mange artikler ud som der er nu.

Question: *Layout af artikler?*

Response: Nej, jeg er ligeglad. Det er blændværk... selvfølgelig kan man sige at noget er flot sat op, men det er ikke noget der siger mig noget.

Question: *Det sidste begreb det er peer review.*

Response: Det har jeg det ambivalent omkring, fordi på den ene side så har man i de artikler jeg har fået ud, så får man sit review tilbage, og så kan man se typisk imellem tre og fem reviewere der er på artiklen, og så kan man se at en til to har ikke fattet en kæft af hvad det handler om, for det de spørger om det skrider til himlen at de ikke ved noget om området. At de måske er specialister indenfor et andet område, og så ved et uheld er de sat til at være reviewere der har med det her at gøre. Jeg har haft een en gang, der skrev "Når du nu siger, at det er neuroner, ved du så det?", det svarer lidt til at sige til en

mekaniker, "Når du nu siger at du har sat fire hjul på, ved du så at det er hjul der sidder på den bil?", så bliver man træt, for det skal man selvfølgelig på en høflig og pæn måde skrive tilbage til manden at ja, det er vi rimelig sikre på, ligesom jeg er rimelig sikker på at der sidder et hovede på skulderen. Men så er der også nogle gange hvor man får et review tilbage, hvor man kan se det virkelig er nogen der ved noget om det, og de kritikpunkter de har, de er godt set, og de højner kvaliteten af det endelige produkt, for de kom lige med noget skarp indsigt, med nye vinkler. Jeg tror at peer review er rigtig godt, så længe det er de rigtige reviewere der er selekteret. Og det er ikke et spørgsmål om at de skal være positive eller negative indstillede for at de er gode. For jeg har da også siddet med en reviewer i den anden ende, hvor jeg har siddet og tænkt at de ikke havde fattet noget, men de giver tommelfingeren op, så fedt nok... men man kan læse ud fra hvad de skriver, at de ikke rigtig ved hvad det handler om, og nogle af de ting de spørger til de er åbenlyse, og nogle af de ting hvor man godt vidste, at her skulle de have sagt et eller andet, dem skøjter de bare henover. Hvor i andre tilfælde har man andre reviewere der er skarpe og sætter fingeren ind hvor der er ømt.

Question: *Så det du siger er, at det er en god idé, men i praksis virker det ikke altid?*

Response: Ikke altid.

Question: *Det er så baseret på dine egne erfaringer. Er det også noget du hører fra andre?*

Response: Det er det, ja. Jeg er en del af en større forskningsgruppe, og der sidder vi af og til og snakker om det der med at revieweren.... der tager man sig til hovedet med nogle af de ting de skriver, men andre gange er de skarpe. Der må man så vende den om og sige at der har det også meget at gøre med det tidsskrift man sender ind til. Altså de virkelig skarpe reviewere de kommer altså fra de tidsskrifter med højest impact. Det skarpeste review jeg har fået det kom fra *Journal of Comparative Neurology*. Jeg fik ikke min artikel ind i det tidsskrift, men da jeg så havde rettet den til ud fra de ting reviewerne sagde og sendte den ind til et lignende tidsskrift, da strøg den lige igennem, men der kunne man også se på nogle af de ting reviewerne sagde, selvom de er anonyme, så kunne man godt regne ud hvem det var, og det var to professorer fra nogle af de rigtig hotte steder de havde fået til at reviewe artiklen, så det var i orden. Men de gav begge to tommelfingeren nedad, fordi de sagde at det ikke var nyt nok til det tidsskrift. Så en reviewer, for at være god, skal ikke nødvendigvis være positiv, men de skal vide hvad de snakker om. Og det er de to bedste reviews jeg nogensinde har fået - de gav begge tommelfingeren ned, men det var virkelig konstruktivt det de kom med, og det var et meget bedre produkt der blev sendt videre til det næste tidsskrift.

Question: *Det var sådan set det jeg havde. Har du mere du vil tilføje?*

Response: Jeg kan ikke lige komme på noget.

A.2.9. Interview reference: M15

Date: Oct 26, 2011

Time: 11:00

Question: *Du har taget to artikler med, og det som jeg godt kunne tænke mig i den sammenhæng er at du fortæller mig lidt om hvorfor du har taget dem med. Jeg har jo bedt dig om at tage nogle med som du vurderede som høj kvalitet, og det er det jeg gerne vil vide, hvad er det, der er høj kvalitet ved de to artikler?*

Response: Jeg har taget de to artikler med, fordi jeg plejer at citere dem, fordi de begge to repræsenterer introduktionen af nye koncepter. Og at de begge to involverer forsøg, hvor man ikke kun laver associationer, men hvor man rent faktisk ændrer på nogle ting, og så ser hvad der sker. Så tilbageændrer man nogle komponenter og ser hvad der sker. Det er nok grunden til at jeg har taget dem med.

Question: *Nye koncepter - vil du prøve at uddybe det en lille smule?*

Response: Altså de handler begge to om et protein jeg arbejder med, som hedder CD36. Og i det ene da viser de CD36's, den der hedder artikel (2), CD36's betydning for udvikling af åreforkalkning. Og det viser de ikke kun ved at associere, hvor man måler på CD36 og så ser på hvor meget åreforkalkning et dyr eller en person har, men de gør det ved at fjerne genet og så se, får du så mindre åreforkalkning, sætte genet ind igen, for de så mere åreforkalkning igen? Og de har nogle passende kontroller for det. Så de laver et passende bevis for at CD36 spiller en vigtig rolle for åreforkalkning. Det er den ene, den anden der er det CD36's rolle for dannelse af blodpropper, altså for påvirkelse af koagulationen, hvor de også går i detaljer og finder ud af at CD36 påvirker blodpladerne og påvirker blodstørkningen ved at binde nogle lipidfraktioner som man ved fra andre studier er associerede med at blodet har større tilbøjelighed til at størkne, og så finder de også hvilke komponenter, hvilke kemiske komponenter inde i de der lipidstrukturer, der faktisk er ansvarlige for det. Og hvor det binder sig på CD36, og det vil sige det er nogle meget dybdegående artikler, som ikke bare er associationspræsenterende.

Question: *Og det, som er nyt ved dem, det er at de gør det på den måde?*

Response: Det der er nyt ved dem er, at den er den første der beviser at CD36 har noget med de to ting at gøre. Ellers har man fundet associationer, eller man har fundet mutationer som har været associeret med den her sygdomstilstand. Men dem her laver en model og så tester de modellen, og dermed beviser det.

Question: *Okay, ja. Er der nogle konkrete karakteristika ved de her artikler som du vil sige, "Det her, det er god kvalitet", ud over det du har nævnt?*

Response: Altså det der er god kvalitet, de har jo så en ordentlig hypotese som er baseret på nogle tidligere studier, så de har altså en velformuleret hypotese, og så har de nogle metoder som rent faktisk undersøger den hypotese man har, undersøger nogle sammenhænge ved at lave et indgreb og se effekten af det, og så med en hel masse relevante kontroller på.

Question: *Hvad med sådan noget som... øhm*

Response: Det er det eksperimentelle jeg snakker om, vil du gerne snakke om konklusionen?

Question: *Nej, jeg synes faktisk det er rigtig spændende, fordi det jo netop er det eksperimentelle, og det lyder også som om det er en lidt anden type forskning end den almindelige, kliniske forskning?*

Response: Ja, det er det også. Altså jeg ville jo gerne at det var mig selv der havde lavet de her artikler. For jeg laver jo rigtig mange af sådan nogle associationsstudier, eller undersøger nogle patienter, jeg kan måle på en markør i en blodårøve, det repræsenterer det her CD36-protein. Her beskriver de det inde i nogle forskellige væv, jeg kan måle det på en blodprøve. Så undersøger jeg så om forskellige sygdomsenheder er associeret med en ændret koncentration af CD36, eller jeg undersøger hvis man nu giver dem nogle tabletter, hvad sker der så? Kan forhøjelsen så gå tilbage? Og så laver jeg nogle hypoteser om dels hvad der repræsenterer det jeg måler, det er nemlig mig der har fundet ud af man kan måle det i plasma, og så nogle hypoteser om hvordan selve patofysiologien, altså sygdomsudviklingen finder sted. Men jeg har jo ikke brugt modeller hvor jeg har modificeret, jeg kommer til at gøre det på et tidspunkt, fordi jeg kommer til at måle på nogle patienter med nogle kendte mutationer i genet, men jeg har ikke lavet sådan nogle helt tilbundsående forsøg, fordi det er for svært for mig.

Question: *De her to studier her, det er vel en slags grundforskning i virkeligheden?*

Response: Ja. Mit er også en lille smule grundforskning. Fordi det kun er mig der kan måle det der protein i plasma, på den måde er alt hvad jeg finder noget nyt, for der er ingen der ved noget om det på forhånd. Men det er jo ikke grund-grundforskning.

Question: *Det jeg sidder og tænker på er noget med det her studiedesign de har brugt her - har jeg forstået det ret, at det var på nogle dyrde gjorde det?*

Response: Ja.

Question: *Ja. Hvis man sådan tænker på evidensniveauer, så sådan noget som dyreforsøg, det rangerer jo faktisk relativt lavt...*

Response: Men det er jo fordi her skaber vi ny viden. Mens hvis det er man skal teste

noget nyt medicin, så duer det selvfølgelig ikke bare at teste det på nogle dyr, eller nogle nye behandlinger, eller noget. Men når man genererer ny viden, og kigger på de helt basale baggrunde, mekanismer, bag den nye viden, så kan dyreforsøg være gode, fordi der kan man gøre noget man ikke kan gøre ved menneskeforsøg. Og så kan man på den måde udbygge sit hypoteseapparat, og så kan man så teste det i det omfang det kan lade sig gøre på mennesker bagefter.

Question: *Ja, og det jeg så gerne vil høre er nemlig, om der er en sammenhæng imellem evidensniveau og kvalitet. Jeg har hørt mange klinikere sige at det er der, men her lyder det egentlig mere som om man sagtens kan lave god forskning, uden at det har et højt evidensniveau?*

Response: Ja, hvis evidensniveau er på den der... men prøv du så at definere evidensniveau.

Question: *Ja, det er så det. Men det er måske også et spørgsmål om at man bruger forskellige evidens..*

Response: ..bruger forskellige termer, ja, det tror jeg. Altså med den her er der et meget højt evidensniveau, fordi man har gjort alle mulige krumspring, for at vise at den sammenhæng eksisterer, som man har en hypotese om er der. Man har haft nogle mus med åreforkalkning, og så har man ligesom fjernet effekten af det gen som producerer CD36, og så har man set, "Bliver åreforkalkningen mindre? Ja, det bliver den", og så har man haft nogle som ikke kan få åreforkalkning, altså uden CD36, og så har man set om de får åreforkalkning, og det gør de ikke. Så har man prøvet at sætte CD36 genet ind i de mus der ikke kan få åreforkalkning, og så se om de så fik åreforkalkning, og det gjorde de så. Man har lavet en hel masse forsøg, som skal bruges til at sige, at CD36 og åreforkalkning har noget med hinanden at gøre. Og så stopper den jo der. Men det er fordi det er et andet type niveau, jeg synes der er meget høj evidens i dem her, det er derfor jeg har valgt dem.

Question: *Er det dét, der er ætiologisk forskning?*

Response: Ja, det kan man sige, bare på et endnu lavere... patofysiologi. Ætiologi er ligesom årsagen til at man får en sygdom, og patofysiologi det er beskrivelsen af selve sygdomsudviklingen. Så man kan ikke sige ud fra det her, at den eneste grund til at man får åreforkalkning det er det her protein, men man kan sige at det spiller en rigtig stor rolle, i hvert fald i en musemodel, og lad os nu teste om det også gør det i mennesker. Det er ligesom på det niveau. Man kan sige enten ætiologi eller patofysiologi, sygdomsudvikling.

Question: *Og der bør man bruge en anden evidensinddeling end hvis det er behandling?*

Response: Ja, det synes jeg, helt klart. Altså jeg synes jo sådan noget her det er meget

større forskning end at putte nogle piller i folk og så se hvad der sker. Meget, meget højere rangerende, meget klogere forskning. Det andet er også nødvendigt, men det her synes jeg det rykker, fordi det flytter nogle paradigmer.

Question: *Ville der så på baggrund af det her senere komme klinisk forskning?*

Response: Ja, nu har jeg jo så på grund af sådan noget her begyndt at interessere mig for CD36, så har jeg fundet ud af man kan måle det i en blodprøve. Her har vi jo taget blodkarrene ud fra de her mus, og kigget på dem, og man har lavet nogle ting man ikke kan lave på mennesker. Nu har jeg så fundet ud af man kan måle det på en blodprøve. Så har jeg så brugt nogle år på at se om jeg så kan finde de her sammenhænge. Hvis det er højt, har det så noget at gøre med hvor meget åreforkalkning vi har. Fordi det med hvor meget åreforkalkning folk har er en klinisk relevant problemstilling, hvis det er højt, har det så noget at gøre med om man får propper, og de andre sammenhænge man kender fra vævet der er beskrevet. Det ser jeg så om jeg kan bruge min blodprøvemærkør til. Det tager lang tid, fordi man også skal lave nogle metoder til det, men hvis det så ender med at forskningsmiljøet [...] accepterede det her, at det var en skidegod mærkør, som man kunne bruge til et eller andet, man opsporede sygdommene tidligere, kunne følge behandlingseffekten, så kan man sige det her, det danner så grundlag for noget der bliver klinisk anvendt 10 år senere, 15 år senere. Og i princippet kan det her også bruges til at lave medicin ud fra, altså selve konceptet kan gøre at man interesserer sig for lige præcis den del af [?] der påvirker blodets størkning, og så kan man lave noget medicin rettet imod den proces, så det er grundlaget for de andre.

Det jeg har lavet ud fra det her, det har jeg jo patenter på, så på den måde er det basal [forskning], for det får man jo ikke på noget som ikke er nyt, og som ikke har et potentiale for at blive anvendt.

Question: *Er det en typisk forskningsproces, det her med at der er nogen der foretager noget grundforskning, og så er der nogen der bygger videre på det, og på et eller andet tidspunkt om 5, 10, 15 år så fører det til en ny behandlingsform, kliniske trials, og så videre?*

Response: Måske fører det til en ny behandlingsform, eller også så dør det ud undervejs. Altså jeg har lavet forskning i mange år, og der er jo altid sådan nogle bølger af hvordan man tror tingene hænger sammen. Det er fordi jeg har lavet sådan noget patofysiologisk forskning, og det har været indenfor type-2 diabetes og åreforkalkning. Så er der jo så f.eks. indenfor årsagen til sukkersyge en bølge i 90'erne hvor alle løb efter den samme bold: det her hænger simpelthen sammen med noget med musklerne, og så var der så nogen der fandt ud af noget med noget betændelse og at de blev tykke, de her mennesker. Så er alle løbet efter den bold, og så nu er der så nogen der tror det har noget at gøre med tarmfloraen, så løber alle efter den bold, så på den måde bliver der jo søsat noget ud fra nogle basale studier og så er der jo en hel masse der laver en hel masse forskning indenfor det. Og det er meget vigtigt i starten at snævre det ind, og så kommer det i virkeligheden til at fylde tre linier i beskrivelsen af den her sygdom når

der er gået 15 år, fordi så er der bygget ny viden ovenpå, som også fylder tre linier. Og sådan bliver der hele tiden bygget ovenpå, og vi bliver klogere. Alle de ting man læser, det der står i pressen med at nu har vi fundet årsagen til det ene og det andet, det er kun i forhold til den viden man har lige for tiden, og den viden ændrer sig hele tiden. Der er ikke nogen sandhed i forskning, men vi er på vej mod sandheden hele tiden. Eller vi søger efter den, men der er altid en hel masse parametre.

Question: *Vil du mene at det er vigtigt, at den kliniske forskning bygger på grundforskning?*

Response: Det kommer an på hvad det er. Hvis det er noget om hvordan man sætter hofter i, f.eks., så ved jeg ikke hvordan man laver grundforskning i det, eller det gør man måske med nogle materialer eller noget, altså det kommer lidt an på hvad det er. Men jeg tror det er uundgåeligt, at det gør det. Men det er bare en proces man ikke ser sammenhængen i, fordi tidsforskydningen er så lang. Jeg tror faktisk altid det gør det, fordi der er nogen der har fået en idé på et tidspunkt, som så enten har ledt til en behandling eller nogle piller eller en operationsteknik eller et eller andet. Og det springer ud af grundforskning, og når man så er oppe i den kliniske forskning og man tester det, så er vejen ned til grunden, den har man glemt.

Question: *Så der er ikke rigtig referencer fra den kliniske forskning til grundforskning?*

Response: Ikke nødvendigvis.

Question: *Og det var der måske nok hvis man fulgte det som en slags slægtstræ?*

Response: Ja, men det gør man jo heller ikke, for så mange referencer må man jo slet ikke have med, når man skriver artikler.

Question: *Nej, det er jo også det.*

Response: Så det forsvinder undervejs, og det der var så stort, der er kun en lillebitte del der kommer, der har en relation til det kliniske projekt der finder sted lang tid senere.

Question: *Ja... Er det muligt at evaluere kvaliteten af forskning? Gerne objektivt...*

Response: Ja, til en vis grad. Det forsøger vi jo at gøre når vi reviewer. Men der ligger man jo nogle forskellige kriterier ind alt efter hvilket tidsskrift det er. Der skal jo mere til at komme i Nature, end der skal til for at komme i Scandinavian Journal of et eller andet. Men der sidder man jo og kigger på; er der en ordentlig hypotese, er der en ordentlig baggrund for at lave det her studie, har man lavet metoderne ordentligt, er det validerede metoder og har man kontroller med, har man lavet ordentlig statistik, og så kommer præsentationen af resultaterne og fortolkningen. Og det er jo også det man bliver bedt om når man reviewer en artikel, i hvert fald i nogle tidsskrifter, at rate de forskellige kategorier, og det er vel et forsøg på at objektivisere kvaliteten. Men jeg

tager da selv hensyn til hvad det nu er for et tidsskrift.

Question: *Hvis så vi snakker om den allerede publicerede forskning?*

Response: Der kan man jo gøre det samme.

Question: *Men det ville simpelthen kræve at man individuelt gik ind og evaluerede?*

Response: Jamen jeg synes hvis man skal lave en ordentlig vurdering vil det kræve at man kan gøre det individuelt. Og ellers kan man jo dele det op i forskellige grupper af forskning; om det nu er sådan nogle associationsstudier eller interventionsstudier eller grundforskning, eller hvad det nu er. Så skal man jo bare bestemme sig for hvad der er vigtigst, og hvordan man skal rate dem, og det kan man jo nok ikke blive enige om. Det kommer nok an på hvem man er og hvor man står. Vil jeg tro. Så kan man jo kigge på citationer, altså hvor mange gange ting er blevet citeret, men det er jo heller ikke en ren parameter, for det kommer an på hvilke områder, om det er et område med stor forskningsaktivitet eller lille forskningsaktivitet, eller publikationsaktivitet.

Question: *Hvad så hvis man kun sammenligner indenfor eet område?*

Response: Så kan man til en vis grad bruge det, men så er det jo også noget med... de gamle mænd citerer hinanden, og hvis man så ikke lige er kammerat med en gammel mand når man kommer med noget nyt, så tager det længere tid at blive citeret. Altså det er helt klart, og det er i meget høj grad tilfældigt hvem folk citerer. Så det viser noget, men ikke det hele. Og man er også tilbøjelig til at citere - hvis man nu har fire artikler som næsten handler om det samme, så tager man jo enten det der var først, eller også tager man det der var i Nature, fordi man tillægger det en højere kvalitet eller en højere evidens.

Question: *Hvis du nu sad i den situation at du skulle give en reference til et eller andet emne, og som du selv siger, så er der grænser for hvor mange referencer vi må give, så hvis man skal være økonomisk med sine referencer, så siger du, at du ville vælge den der havde været i f.eks. Nature?*

Response: Nu er det jo helt hypotetisk at der står det samme i dem, det gør der jo aldrig. Så det skal være den, der læner sig mest op ad det jeg vil, f.eks. forsøge at vise folk, at der er en grund til at jeg laver det her studie, f.eks. i baggrunden eller i diskussionen. Så ville jeg kigge på hvilken der var mest relevant og til en hvis grad hvilken reference der var tidligst også, hvis det nu var noget banebrydende, hvor der var tre grupper der var kommet noget nær. Og så ville jeg nok ikke tage det fra Turkeys Journal of et eller andet, jeg ville meget hellere tage et fra højere rangerende tidsskrift.

Question: *Hvorfor ville du gøre det?*

Response: Det er fordi jeg forventer at reviewerne på de tidsskrifter har... det borger for kvaliteten at det er publiceret i et højere rangerende tidsskrift.

Question: *Nu må du rette mig hvis jeg tager fejl, men det lyder som om det du siger er, at hvis referencen er til Nature frem for et eller andet lidt mere obskurt, så er det et bedre argument for din påstand?*

Response: Altså jeg vil tage Nature frem for den tyrkiske journal, for hvis det der stod i den, man kan sige, jeg læser selvfølgelig referencen, men jeg sidder jo ikke og bedømmer den, jeg kan jo ikke sidde og bedømme alle de her artikler jeg refererer til. Jeg har noget indhold som kan bruges til at understøtte det jeg skriver. Og hvis det nu var lige så mega-fantastisk det der stod i Turkish Journal artiklen, så ville de jo have sendt det til Nature eller et andet højere rangerende tidsskrift. Altså, der er jo ingen der starter med at sende den ind til... jeg starter jo ikke med at sende ind til Ugeskriftet, jeg starter for højt, altid, fordi jeg synes det er skidevigtigt det jeg skriver. Så derfor, på den måde, så lægger jeg en kvalitetsstempling [på tidsskriftet]. Hvis det nu var to diabetestidsskrifter, der er et europæisk og et amerikansk, det amerikanske rangerer højest og det er fordi, sådan er det med amerikanske tidsskrifter, så ville jeg ikke nødvendigvis tage det amerikanske fordi det har en impact factor på 9, og det europæiske har på 6.5. Så vil jeg meget mere gå ned i detaljen, eller i det detaljerede indhold. Så det var så yderpunkter.

Question: *Så det er en kombination af indholdet i artiklerne, og at man kan forvente at hvis man kommer med en reference til Nature, så vil læseren også i højere grad forstå at det er en god reference?*

Response: Det er jo en garanti for en selv, at man referer til noget, som ikke er noget skrammel. Jeg tænker, at det jo ikke er ret mange læsere, der kigger på de referencer.

Question: *Så det er egentlig mest for sig selv?*

Response: Ja, det tror jeg der var i gamle dage, altså det er jo fordi man skal have den akademiske diskussion. Men det tror jeg meget mere var i gamle dage, at man gik ind og læste de andre. Hvis man er helt ny indenfor et område, så vil man gå ind og kigge på referencerne, måske. Men nu om dage er der så mange publikationer, at der også kommer en højere grad af tilfældighed ind i hvad det nu er man referer til. Så hvis jeg nu skulle ind i et nyt emne, så ville jeg da søge i PubMed og så kigge på det selv. Så selve det med referencerne bliver man nødt til at have for at bibeholde den akademiske, altså at bibeholde det at skrive en artikel som en akademisk disciplin, og der skal selvfølgelig være noget belæg for at det ikke er fuldstændig løsrevet fra resten af verden. Men referencelisten betyder ikke så meget mere, synes jeg, som den gjorde engang. Det gør den når man skriver oversigtsartikel, men i oversigtsartikler kan man jo heller ikke engang have alle, selv dengang jeg startede med det der protein for 10 år siden, da tror jeg der var 50 referencer om det i det hele taget, og nu er der 25,000 eller

sådan noget, og det er der jo ingen mennesker der kan følge med i.

Question: *Jeg synes det er meget interessant det du siger med tilfældighed. Vil du så også mene at... jeg må hellere lige spørge om noget andet først... hvis en artikel er højt citeret, er det så også en indikator for at den har en høj kvalitet, eller er det fuldstændig tilfældigt? Eller er det bare de gamle mænd der sidder og citerer hinanden? Eller?*

Response: Jeg tror der er flere komponenter. Det har en kvalitet, selvfølgelig, hvis det er højt citeret. Og den kvalitet kan jo være at det er nyt, eller at det er et stort studie med mange deltagere, eller sådan noget, så der er nogle faktorer der gør at den bliver højt citeret. Så betyder det også noget om det er en af gruerne indenfor området, og så var der en ting mere som jeg lige skal komme i tanker om, jo, og så er det jo tit at folk citerer de samme som andre har citeret. Hvis man nu sidder og skriver, så har man sådan et... altså fordi man ikke kan følge med i al litteraturen indenfor ens område, det kan jeg i hvert fald ikke, og så når man sidder og skriver om et område man ikke har så meget forstand på, så kigger man lidt på nogle artikler eller oversigtsartikler og så vil jeg være mere tilbøjelig til at se mere på de artikler som andre har citeret, hvad er det egentlig de skriver der, i stedet for at skulle kigge på 100 artikler og så vælge dem ud der var bedst.

Question: *Fordi det jeg så synes det var lidt interessant ved at det var lidt tilfældigt hvem det så var der kom på referencelisten...*

Response: ..men det er jo tilfældigt, ud fra det jeg siger her.

Question: *Men vil det så også betyde at hvis man får mange citationer, så skiller man sig i højere grad ud end før, hvor der ikke var ligeså meget information?*

Response: Det ved jeg ikke, det kan man jo godt udlede. Det tror jeg, jeg skulle tænke længe over for at blive helt enig med mig selv. Men det kan man jo ikke helt udelukke, det er jo ikke forkert det du siger, ud fra det jeg siger. Og så er der jo det med citationerne, at de holder op igen. En artikel lever i et vist antal år, og de allerfleste dør ud fordi dem der skriver har glemt at det var fantastisk for fem år siden. Jeg har også tidligere lavet noget der var helt nyt, og så bliver jeg vældig citeret i nogle år. Hvis jeg kigger på dem, der er mest citeret, så er det dem hvor der er nye ting i, og så nogen hvor der er bred appeal. Sådan noget med unge mænd der træner, eller sådan noget. Det kan fitte ind i mange sammenhænge, mens meget af det andet jeg har skrevet, det er sådan nogle småbitte molekyler jeg har arbejdet med, og så er det så nyt, fordi det har jeg lige fundet, men så mister det så appealen fordi på et tidspunkt bliver det bare er generel knowledge. Det der er med det protein jeg har fundet i blodet, det har man så skrevet lange... dem der har citeret mig, jeg tror den første af dem var fra '06, og den tror jeg er citeret nogen-og-fyrre gange, og det er jo fint nok. Og dem der har gjort det, det er jo fordi de skulle skrive, at vi har det her nye molekyle og sådan og sådan og sådan, mens

nu er folk allerede begyndt at skrive artikler, hvor de ligesom tager det for givet at det findes.

Question: *Okay, så de diskuterer overhovedet ikke hvor det kommer fra?*

Response: Man diskuterer slet ikke “kan det nu passe”, eller “hvad er der nu af evidens for at det her molekyle findes”, det gjorde de til at starte med, så refererede de til mig, nu er det så bare sådan at det molekyle jeg har, det er ligesom linket op med det der findes inde i vævene, det er ligesom fact. Så det betyder også noget for antal citationer, så behøver man jo ikke citere det mere.

Question: *Når vi nu snakker om citationer, ved du så hvor meget de her to er citeret?*

Response: Nej, men jeg vil tro, hvis jeg skulle gætte, over 100 begge to. De er fra, arh, '07, den er nok ikke over 100, jeg vil tro den er omkring 100, måske 80.

Question: *Okay. Det sidste jeg har er en slags brainstorm, hvor jeg vil nævne nogle begreber, og så vil jeg meget gerne have du svarer det første du tænker på, selvfølgelig i konteksten forskningskvalitet. Det første begreb det er ny viden.*

Response: Ny viden... jamen der tænker jeg på grundforskning.

Question: *Er det vigtigt?*

Response: Ja...

Question: *...det er måske også et lidt ledende spørgsmål.*

Response: Jamen det er jo supervigtigt. Det er det som politikerne ikke har forstået, men som de måske begynder at forstå igen på et tidspunkt, det der med at styre forskning, og give forskningsmidler til på forhånd definerede områder. Hvis man gør det, så får man nogle centre, der bliver rigtig store i verden, men man mangler de små miljøer, der går og... altså, man finder jo ikke på nye ting, når man går i sådan en stor boble, og arbejder omkring det samme koncept. De nye ting finder man hvis man får en lille boble, og så snakker man måske med nogle andre i en anden lille boble, så kommer der ny viden. Så derfor er det vigtigt at holde en hel masse små grundforskningstiljægere igang, så de der underlige interaktioner, når de nu møder een, de ikke plejer at drikke kaffe med, så er det, de nye idéer opstår. Så er den idé ny i et stykke tid, så bliver den ligesom groet fast i koncepterne indenfor den, og så sker der egentlig ikke ret meget andet, så skal der være nogle andre, forskellige, der snakker sammen, som kan lave nogle nye koncepter igen. Ellers så kommer vi altså ikke fremad, derfor er grundforskning vigtigt. Så er det jo også basis for hvad der kan ske sidenhen, det laver ligesom fundamentet for anvendt forskning sidenhen.

Question: *Hvad med layout af artikler?*

Response: Det synes jeg også er vigtigt. Men det er jo ikke sådan at jeg kun læser dem der er pæne. Men jeg tænker da på om artikler... altså noget de ser nærmest stikkende ud, fordi de bruger nogle forkerte bogstaver, eller det er for kompakt.

Question: *Men kunne du finde på at undlade at læse en artikel hvis den simpelthen var "grim"?*

Response: Hvis der f.eks. slet ikke var nogen figurer i, for eksempel... nej, det tror jeg ikke. Men jeg synes da det er meget forskelligt.

Question: *Hvad med sådan noget som sproget, er det vigtigt?*

Response: Det er enormt vigtigt. Det kan være døden at læse en der er skrevet dårligt. Og det kan jo være svært, det kan være svært at skrive ordentligt. Også fordi man også der jo har restriktioner, også hvor meget, man har jo så meget man vil fortælle, og så skal det lige fylde mindre og mindre.

Question: *Hvad med impact?*

Response: Så tænker jeg på hvilken betydning det har for fremtid forskning, hvis du skulle have det første jeg tænkte på.

Question: *Tværfaglighed?*

Response: Tværfaglighed, da tænker jeg mest på klinisk forskning, der er tværfaglig. Men så tænker jeg også på den tværfaglighed jeg forklarede om før, at tværfaglighed kan give nye koncepter.

Question: *Altså det med de små bobler?*

Response: Ja, de gange hvor jeg har lavet noget der var genialt, det kan man jo ikke sige, men noget som ikke bare var almindelig gadagung-gadagung, det har været sådan når jeg har mødt nogen der lavede noget helt andet, og vi har hørt hvad hinanden har sagt og så, "Nåh ja, sådan kunne det måske også være". Jeg mener selv det er vældig vigtigt.

Question: *Kunne det være noget man også kunne opnå ved at opsøge artikler fra andre fagområder end ens eget?*

Response: Eller bare emneområder, det behøver ikke engang være fagområder. Fordi fagområder så skulle man begynde at læse ingeniørfaglige artikler, f.eks. Det kan bare være nogle andre tidsskrifter indenfor lægevidenskaben, det gør jeg nogle gange, hvis jeg f.eks. tilfældigvis kommer til at se - nu får jeg mange idéer i forvejen - men jeg får

altid en masse idéer når det er fordi... så kunne man lige tænke, at det kunne være at det kunne være på samme måde indenfor ens eget felt, eller man kunne overføre nogle af begreberne eller nogle af metoderne eller sådan noget. Så det tror jeg faktisk, hvis man skal holde det igang, så er det vigtigt.

Question: *Hvordan opdager du de artikler?*

Response: Det er Google, jeg er verdensmester i Google. Jeg googler enormt meget, så googler jeg f.eks. efter noget protein eller et eller andet, og så er der noget mere, og så kører jeg videre ud af den, og så ser jeg noget som måske kunne være spændende, som jeg ikke ville have læst hvis jeg bare søgte artikler som jeg plejede indenfor mit felt. Og det tror jeg da er en god idé, udviklingsmæssigt.

Question: *Det sidste jeg har, det har du måske svaret lidt på allerede. Det er peer review.*

Response: Ja.

Question: *Har du mere at sige til det end det du allerede har, i forhold til forskningskvalitet?*

Response: I forhold til kvaliteten? Nej, der har jeg sagt at man har de der kriterier og sidder og bedømmer ud fra hvilket tidsskrift man nu er sat til at bedømme for. Og så er der jo en vis grad af farvning også; altså peer review er jo ikke objektivt, slet ikke. Man får jo simpelthen de mest lamme kommentarer nogle gange, på sine egne artikler, hvor man tænker, altså når man er blevet færdig med at være sur, at det kan de jo ikke have læst ordentligt, det har de jo slet ikke forstået, og så må man jo se om man kan skrive det bedre. Eller også bliver der sprøjtet sådan nogle, at man skulle lige lave noget som ender med at være umuligt, altså sådan nogle blokader ud, i forhold til at kunne komme til at publicere. Jeg synes ikke nødvendigvis det er objektivt, altså jeg synes der er meget... der er følelser i det også. Så skal man jo i vid udstrækning, hvis der nu er nogen man tror der er uenige med en, i ens koncepter, så skal man benytte sig af det med at skrive hvem man ikke vil have som reviewer. Fordi det tror jeg faktisk de respekterer. Det har jeg selv ikke gjort i lang tid, fordi jeg synes det virkede mistænksomt på en eller anden måde, jeg mistænkeliggjorde mig selv, men jeg tror egentlig det bliver taget meget mere straight.

Question: *Det er simpelthen bare et spørgsmål om at man ved der er nogen der repræsenterer et andet standpunkt, så for at undgå de skal lægge blokader ud, så?*

Response: Som ikke er enige i ens koncept, ja, og så samtidig også det med at man kan ønske sig reviewere. Det har jeg også syntes var noget mærkeligt noget, men det er jeg også begyndt at gøre, og det er jo faktisk sådan at dem man tror vil være ens største konkurrenter, er dem der arbejder med det samme emneområde, sådan har jeg det i hvert fald, sådan er det jo slet ikke. Det er jo fordi de interesserer sig ligeså glødende for det som man selv gør, så dem skal man bare sige, man skal tage sine tre værste konkurrenter

i forhold til at være først og lade dem kigge på det, de ved jo hvad det handler om, og de er interesserede i det, i stedet for at det bare er sådan noget venstrehånds-noget. Men det er jo forskelligt, nogle gange når jeg selv peer reviewer, så er det jo heller ikke objektivt, altså hvis de nu har skrevet helt fænomenalt ringe, så læser jeg den ikke ligeså ordentligt som hvis nogen har skrevet godt, fordi jeg simpelthen går død i det og ikke gider.

Question: *Det var sådan set det jeg havde så. Du skal have mange tak.*

A.2.10. Interview reference: M16

Date: Oct 20, 2011

Time: 09:00

Question: *Du har taget de her to artikler med, og dem har jeg bedt dig tage med på baggrund af at du vurderede dem til at være høj kvalitet. Vil du fortælle mig noget om hvad det er ved de her to artikler du vurderer som høj kvalitet?*

Response: Jamen det er to artikler der handler lidt om det samme. De har været med til at revolutionere - den første artikel her (1) fra 2001, har været med til at revolutionere den måde vi behandler septiske patienter på, og det er egentlig derfor jeg vurderer den til høj kvalitet, fordi den har umiddelbar betydning for den måde vi handler i det daglige. Altså vor man kan overføre det teoretiske til det praktiske. Det er for mig at se høj kvalitet. Jeg er jo ikke den store statistiker, så på den måde har jeg jo ikke kunne gennemarbejde dem, men jeg går ud fra at den side også er i orden siden de er blevet implementeret så hurtigt. Så noget der umiddelbart får betydning for det jeg arbejder med i det daglige, det vurderer jeg som høj kvalitet. At den her impact.

Question: *Så skal jeg lige høre.... er du primært kliniker?*

Response: Ja, det er jeg. Jeg har sat kryds ved forskning, der er ikke noget jeg har publiceret endnu, men jeg er igang med at hjælpe med et forskningsprojektovre på Øre-, Næse-, Hals[kirurgisk afdeling], hvor jeg forventer at skulle igang med det teoretiske også. Lige nu er det det praktiske.

Question: *Hvad med den anden artikel?*

Response: Den anden artikel, det er så en oversigtsartikel, der er noget nyere, fra 2010, der ligesom samler op på hvad man rent faktisk gør i det daglige i forhold til hvad man fandt ud af i den første artikel. D.v.s. der er små ti år imellem dem, og den sidste artikel er mere en oversigtsartikel, hvor man opsummerer hvad man rent faktisk anvender nu om dage af de ting man fandt ud af i 2001. Man kan diskutere om det er en god artikel - Jo, den er god fordi den opfylder det der er formålet med den, den giver et overblik over

behandlingsstrategierne ved sepsis. Og det synes jeg egentlig den gør på en udmærket måde. Om de har nok artikler med, det har jeg ikke været i dybden med.

Question: *Er der noget du vil fremhæve som særlige karakteristika, som gør at det er høj kvalitet, eller er det bare det, at den er anvendelig.*

Response: Umiddelbart det med at den er anvendelig. Hvis du går ned og kigger på artiklen, så er det sådan noget med om de har mange med, om den er velplanlagt. Det er jo klart, det er en kvalitet ved artiklen. Og så lader man sig jo også rive lidt med, jeg tror den her Rivers-studiet er et af de første og mest citerede studier indenfor vores felt i hvert fald. Og det er klart at der måske også er lidt Kejsers nye Klæder ved det, hvis der først er nogen der starter med at citere den, så vil alle citere den. Om det er berettiget, det skal jeg ikke kunne sige, der tror jeg også man skal være mere hård i det statistiske.

Question: *Hvad med sådan noget som studiedesign? Altså sådan noget som evidensniveau?*

Response: Altså umiddelbart så er det et randomiseret, kontrolleret klinisk studie. Så det er jo noget af det ultimative man kan få, synes jeg.

Question: *Er det aførende?*

Response: Ja, det synes jeg det er, i hvert fald i forhold til direkte at kunne overføre det til behandling i klinikken. At man er sikker på at der ikke er noget selektionsbias eller noget, det er med til at løfte kvaliteten af studiet.

Question: *I hvor høj grad bruger du forskningsartikler i dit kliniske arbejde, sammenlignet med f.eks. opslagsværker?*

Response: Det svinger noget. Hvis jeg står med et konkret klinisk problem, så bruger jeg opslagsværker. Heldigvis da online-versioner, så det er nyeste version, hvor der er folk der har sat sig ind i den litteratur der nu er tilgængelig. I forhold til undervisning, som vi jo også laver en del af, så er det friske artikler vi søger. Men i en klinisk problemstilling, for at spare tid, da er det lettere at gå ind og slå op i et online opslagsværk, for så er der nogen der har tænkt tankerne for een og samlet de relevante artikler.

Question: *Ville man kunne få noget ud af i højere grad at bruge forskningsartikler, hvis man nu havde tiden til det?*

Response: Ja, selvfølgelig får du jo den daglige træning i artikellæsning, så på den måde vil du få noget ud af det. Plus at du kan danne dig din egen mening. I opslagsværkerne er det jo dem der samler det, og deres fortolkning af resultaterne, der er jo ikke noget der er sort og hvidt, det er der jo sjældent der er. Så jo, på den måde kunne det være en fordel at man selv kunne gå ind og vægte de forskellige studier, og så selv drage sine

konklusioner derudfra. Mange af vores opslagsværker er jo også amerikanske, og nogle gange er det jo nogle lidt andre forhold de har derovre, nogle lidt andre patientkategorier og så videre. Så der kunne det måske være en fordel at gå ind og tolke på europæiske eller skandinaviske studier, der har man ikke mulighed for hvis man bare går ind og tager opslagsværkerne.

Question: *Ved du om der er forskel på de forskellige læger rundt omkring, I er en ret stor afdeling, er der forskel på f.eks. ancienniteten eller forskningsaktiviteten i forhold til om man bruger opslagsværker eller forskningsartikler i det kliniske?*

Response: Altså jeg tror umiddelbart at mange af os der er unge i faget vi bruger mere opslagsværker, lidt mere ukritisk, end den ældre generation gør. Både fordi de selvfølgelig mere eller mindre kan det, der står i opslagsværkerne i forvejen, det der er basalt, som vi andre skal ind og læse på, det kan de selvfølgelig. Og så begynder de jo at få noget overskud til at kunne bygge videre og danne deres egen mening, holde øje med det sidste nye osv. Så der er nok en tendens til, at det primært er os yngre der bruger opslagsværker og de ældre går ind og søger mere specifikt i litteraturen. Heldigvis er det den vej, for man kunne også have en fordom om at nogle af de ældre måske slet ikke ville holde sig opdateret på artikler osv.

Question: *Det er i hvert fald en fordom jeg har hørt, så...*

Response: Det er ikke mit indtryk, umiddelbart.

Question: *Hvis vi nu skal prøve at snakke om forskningskvalitet helt generelt, har du nogle tanker om det? Vi har lige kigget på to artikler, meget specifikt, men hvad med mere generelt?*

Response: Jeg vil stadig mene det er en kvalitet hvis resultaterne er umiddelbart realiserbare, altså noget der får en konsekvens, jeg synes der er meget meritforskning.

Question: *Hvad tænker du på når du siger det?*

Response: Hvor folk laver forskning for at få stillinger eller for at komme frem. Og meget forskning ender op med at det ikke flytter noget, sådan lidt gymnasie-opgaveagtigt, at man slutter af med at sige, at der er brug for flere undersøgelser. Dvs. vi er egentlig ikke kommet et skridt videre, så jeg synes kvalitet det er klart når man får nogle resultater, der kan ændre på en behandling eller konstaterer at den behandling vi allerede udfører er den optimale. Det er selvfølgelig mere med formålet med forskningen. Kvaliteten ligger jo også i det, at hvis studiet ikke er designet nok, ikke har nok med, ikke inkluderer nok, eller ikke har statistisk power nok til at implementere en ny behandling, så på den måde skal det selvfølgelig også være studier af en vis størrelse og kvalitet for at man kan gøre det.

Question: *Hvad med sådan noget som grundforskning?*

Response: Jamen det er der jo bestemt også brug for. Jeg er mest kliniker, og det bliver jeg ved med at være, selvom forskning også interesserer mig. Så grundforskning er ikke det jeg vil komme til at beskæftige mig mest med, og jeg ved godt at der kan det være sværere umiddelbart at se resultater, men det er bestemt vigtigt.

Question: *Så som kliniker er grundforskning ikke så vigtig i det daglige?*

Response: Den er lidt mere fjern, på en måde, og den er selvfølgelig akademisk interessant og vigtig, men når det ikke er det der sådan lige indenfor en overskuelig tid får betydning for den måde vi behandler patienterne på så er det ikke der jeg vælger at lægge min tid.

Question: *Er det muligt at evaluere kvaliteten af forskningen?*

Response: Ja, det er det jo. Men igen synes jeg det kræver lidt større statistisk indsigt at kunne sæbe en artikel og finde de svage punkter. Det er klart, alle kan jo se på hvor mange, der er inkluderede osv. men sådan virkelig at skulle gå til en artikel på den måde, det kræver en lidt større indsigt.

Question: *Tænker du det ville være muligt at have en slags mål for kvaliteten?*

Response: [hesitation]

Question: *Det er selvfølgelig altid svært at måle kvalitet...*

Response: Det er svært at sige... nu er det nærliggende igen at sige, at der skulle kunne komme en konklusion ud af det, men det er jo ikke altid sikkert, og nu nævner du jo selv grundforskning, og det er jo langt fra sikkert man kan det. Der er sikkert mange der vil gøre det også fra politisk hånd, for så var det mere målbart. Hvad man får for pengene.

Question: *Det er der jo mange, der er interesserede i.*

Response: Ja, men det er nok sværere end som så.

Question: *Der er ikke lige noget du tænker på man kunne se på, nogle kriterier der skulle være opfyldt, eller?*

Response: Nej, det kommer an på... det er jo power-beregninger hvor mange der skal inkluderes for at få signifikante data, så det er vel afhængig af hvad man kigger på.

Question: *Okay. Vi har snakket en lille smule om de her studietyper, f.eks. randomiserede forsøg. Er det dét at et forsøg er dobbeltblindet, randomiseret og kontrolleret osv. der er afgørende for at det er kvalitet eller er det sammenhængen imellem hvad det er for en problemstilling man har og så studietypen man bruger, der er kvaliteten?*

Response: Jeg vil stadig sige det er individuelt hvilket studie der skal til for at undersøge en konkret problemstilling, men sådan som basis kan man sige, vil det altid være en kvalitet med det prospektive studie, hvor man randomiserer folk, at man ikke kigger på det retrospektivt, fordi man så har mulighed for at standardisere forholdene så meget som muligt. Der vil altid være færre ting man ikke kan korrigere for, hvis man laver det prospektivt. Så det er jo det optimale, og det er de fleste også enige i, når det kan lade sig gøre. Meget forskning indenfor vores felt er sådan akut forskning, så der er det ikke altid muligt, der bliver det noget med at kigge tilbage, retrospektivt, bl.a. også fordi man ikke kan få patienternes tilladelse altid. Der er problemstillinger der.

Question: *Er der sammenhæng imellem forskningsartikler og den forskning, der ligger bagved, i forhold til kvalitet, altså kræver en god artikel et godt studie, eller kan man også skrive en god artikel på baggrund af et fesent studie?*

Response: Altså artiklen kan selvfølgelig være velskrevet osv., men jeg synes der skulle lige et godt studie bag en god artikel, der synes jeg der er en klar sammenhæng. Det er jo facts, man kan jo ikke omskrive dem, eller få dem til at se bedre ud i artiklen. Det kan man sikkert godt, men det ændrer ikke på grundstoffet.

Question: *Har du tilfældigvis undersøgt hvor højt citeret de her to artikler er?*

Response: Nej, det har jeg ikke. Men samtlige kurser jeg har været på med behandling af sepsis der er det det her River-studie der bliver citeret. Hvor mange det lige er, det ved jeg ikke, men det må være højt, højt citeret. Det er et som alle har hørt om, det er sådan meget mainstream-studie indenfor vores fag, så det må være et af de mest citerede.

Question: *Hvad med sådan noget som tidsskrifter, har du hørt om impact factor?*

Response: Ja.

Question: *Er det noget du tænker over når du ser på en artikel, hvor det er udgivet henne, og hvad deres impact factor er?*

Response: Nej, det gør jeg faktisk ikke. Det synes jeg ikke rigtig jeg har haft grund til. Det er måske noget man ville se på hvis man selv skulle til at citere noget fra nogle artikler, så ville man selvfølgelig tage noget fra de tidsskrifter, der har størst impact factor. Men jeg synes ikke umiddelbart, nej, når jeg søger på det er det ikke noget jeg lægger mærke til, nej.

Question: *Mener du der er en sammenhæng imellem tidsskrifters impact factor og kvaliteten af de artikler der er deri?*

Response: Det kan der selvfølgelig være, men det tror jeg ikke umiddelbart der er. Nu har jeg selv hørt fra andre hvor svært det er at få optaget artikler rundt omkring, selvom

det egentlig kan være en okay kvalitet, så jeg synes ikke umiddelbart der er, jeg tror mere det er en snob-effekt.

qqSå har jeg nogle begreber som jeg gerne vil høre din umiddelbare reaktion på. Du må selv bestemme hvor meget du vil sige om hver enkelt, men det må meget gerne være i den kontekst der er forskningskvalitet. Det første begreb det er ny viden. **Response:** Ja. Og der vil du have mig til at sige?

Question: *Bare det første du tænker på.*

Response: Jamen i forhold til det her, når vi snakker om forskningskvalitet, det vil være at læse de nyeste artikler, og netop ikke opslagsværkerne. Hvis det er den helt nyeste viden, så er det artiklerne man skal have.

Question: *Så de gode artikler er dem der også har den nye viden, eller?*

Response: Ja, det vil jeg sige, og man kigger da også på hvis man har en klinisk problemstilling og skal ind og søge artikler på det, så starter jeg da med at læse de nyeste, selvom der måske er nogle af sværvægterne der er ti år gamle. Men alle de nye refererer jo også til de tidligere artikler. Så jo, jeg vil altid læse de nyeste. Og er der oversigtsartikler, så er det jo rigtig godt.

Question: *Så det er faktisk en kvalitet i sig selv, at de er oversigtsartikler?*

Response: Ja, det synes jeg de er, for så er der andre der har samlet problemstillingen, og beskriver det mere klinisk anvendeligt.

Question: *Er der forskel på typen af oversigtsartikler i forhold til deres kvalitet? Altså, jeg tænker på systematisk review versus ikke-systematiske, meta-analyser og den slags?*

Response: Ja, det er der formentlig, det har jeg ikke indsigt nok til at gå ind i.

Question: *Hvad med tværfaglighed?*

Response: Min umiddelbare tanke i forhold til forskningskvalitet... det ved jeg ikke lige hvad jeg skal sige til. Så skulle det være tværfaglighed imellem forskellige specialer, der er man nok tilbøjelig til at springe uden om de artikler der ikke lige er publiceret i et tidsskrift for anæstesi, altså der hvor jeg er, der kan jo sagtens være artikler fra andre specialer. Lungemedicinerne kunne have skrevet om sepsis, f.eks., dem er man nok tilbøjelig til at guide udenom og så holde sig til dem der er publiceret i speciale-tidsskrifter.

Question: *Hvad med sådan noget som layout af artikler?*

Response: Det har meget at gøre med læsbarheden, og det har meget at gøre med hvor

appetitlige de er, men jeg vil ikke sige det er afgørende for kvaliteten af artiklen som sådan. Men det er da klart at man ved mange artikler kunne gøre mere ud af at gøre det læsbart og masser af faktabokse osv. Det gør det mere appetitligt at gå til og mere letlæst. Og det har man nok forsømt mange steder, synes jeg, hvor det bare er spalte op og spalte ned uden mellemrum eller noget som helst.

Question: *Peer review?*

Response: Ja... det burde det mest jo være, af det vi beskæftiger os med, så det er ikke noget jeg fokuserer på som sådan. Hvis ikke de nævner det alle steder længere så tager jeg det som en forudsætning.

Question: *Er det en slags kvalitetsindikator, at det er blevet peer reviewet?*

Response: Ja, det synes jeg, det må man tage det som.

Question: *Som minimum eller som?*

Response: Ja, det burde være som minimum. Jeg tror også det er derfor de ikke nævner det alle steder. For ellers er det jo sådan en floskel som de kunne smide på næsten alt.

Question: *Hvad med sådan noget som citationer, er der en sammenhæng imellem dem og kvalitet?*

Response: Tænker du på antal?

Question: *Ja, for eksempel.*

Response: Ja, til dels. Igen er der måske også nogle mode-studier som folk gerne vil citere fra, men som udgangspunkt er det da en kvalitet hvis en artikel er citeret mange gange.

Question: *Jamen det er sådan set det jeg har, medmindre der er noget du sidder og brænder inde med, noget du gerne vil sige?*

Response: Nej, ikke umiddelbart.

Question: *Så vil jeg sige tak.*

A.2.11. Interview reference: M18

Date: Oct 18, 2011

Time: 13:00

NB! The interviewee in this interview spoke a mixture of Danish and Swedish. The responses of the interviewee have been translated to Danish during the transcription.

Question: *Velkommen til. Du har taget nogle artikler med, og dem vil jeg meget gerne have at du fortæller mig lidt om, hvorfor du vurderer dem til at være høj kvalitet.*

Response: Jeg ved ikke om den her artikel (1) siger præcis hvad der er forskningskvalitet. Jeg arbejder selv ofte i forhold til disse kriterier når jeg skriver en artikel, men forskning er så meget mere end artikelskrivning. Så derfor har jeg [...] altså lidt grundprincipper i hvad der er kvalitet i forskning. Og det er at man tænker sig nøje om, hvad er lavet, og det tror jeg er et stort dilemma i dag. Man finder på nogle ting og så går man lige på “what I’d like to do” og så videre... de her spørgsmål, i stedet for “what has been done before?”. Og hvis man gør det, og det fremgår af (1), “What does an editor look for?”, han vil have nye data. i dag produceres en masse sager som ikke påviser noget nyt, det er mere kvalitetssikring og det er hurtig forskning, og man læser dem som referee, og man ser der er mange ting der kommer ind, som ikke er originelle, som ikke er nyt, og hvorfor ikke? Fordi man ikke har læst litteraturen så det er en *circulus vitiosus*, så man kan godt sætte sig ind i en editors eller reviewers opfattelse, for de får jo ganske mange artikler ind. Og så kan man jo spørge, hvad skal man gøre for at løfte dette? For det er jo tydeligt et kvalitetsmoment, der ikke er på plads i dag, og der er min indstilling at “no research is better than bad research” og “bad research” det er præcis dette; intet originelt, bare lidt diffuse ting. Og så kan man gå videre og spørge til årsagen, og det kan jo være tidsaspekter eller helt enkelt at man ikke er tilstrækkeligt stringent i sin tankegange. Hvad er det som er nyt? Hvad er det vi i virkeligheden ikke ved? Og så kan du desuden gå yderligere og spørge, hvad skal du gøre for at komme frem til bedre forskning? Og der er mit forslag at vejledere - når man ønsker den type undersøgelse, man er vejleder eller senior - giver input til idéer eller får det i en plan, at man der både som vejleder og som kandidat, forskerkandidat, laver systematiske reviews. Det tror jeg vil være en god begyndelse for at fokusere på hvad konkret ikke er lavet. Og om der findes et systematisk review om dette emne ved jeg ikke om det er gjort.

Question: *Det er det ikke.*

Response: Jamen så anbefaler jeg det. Og det er sikkert meget vanskeligt.

Question: *Problemet er nok, at der er skrevet så lidt om det, at du ikke ville kunne få et systematisk review ud af det.*

Response: Ja, jeg kender ikke så meget til det, så jeg kan ikke give dig nogle tips om det. Men det kan din vejleder gøre. Jeg kan se at der findes... den anden artikel (2) er en artikel der er lidt bredere, det er et systematisk review, men den er ikke rigtig i følge de regler og krav man har på større systematiske reviews, men det er tre personer som har lavet et fint systematisk review. Den er hurtig til at fokusere, nemlig at der er skrevet meget, man bruger brede søgeord og siden afsætter man næsten et halvt år, tre-

fire personer, gransker litteraturen og ser hvad der er evidens, hvad har man for evidens. For et systematisk review skal jo altid afsluttes med en evidensgradering.

Question: *Ihvertfald indenfor de fag hvor det er muligt, ja.*

Response: Ja, jeg tror da at man videnskabeligt kan applikere tanken på hvad er plus og hvad er minus, hvad er uklart, det vil jeg tro kan passe på de fleste kvalitative data. Men der må man naturligvis også trænes i evidensbasering, så ifølge mine tanker, så tror jeg at enhver som sætter igang - især klinisk forskning - for den har så mange problemer altså "inborn problems", med tanke på patienter, counfounders. At klinisk forskning starter med systematisk review og evidensgradering det er kundskab. Så man sætter sig ned og der kan det blive sådan, som man til tider ser, at man på to-tre år, 300 artikler som siger det ene og 300 artikler der siger det andet, og når du evidensgraderer, så ender du med en eller to, og selv de er contradictory research. Det betyder at... ja hvad betyder det?

Question: *Det betyder, at du ikke kan stole på hverken den ene eller den anden.*

Response: Ja, men hvad betyder det for kvaliteten af forskning? Hvis du havde gjort det i dit emne, eller X emne, og du gør det og du går hjem og tænker over, "nu har jeg på 2 år fundet 600 artikler, 50/50 siger, ja, det skal du gøre, de andre nej, det skal du ikke, og så evidensgraderer du, og finder at evidensen er lav på de fleste. Så gør du hvad for at få bedre kvalitet?"

Question: *Jamen hvis du virkelig har den situation, så har du jo egentlig 600 artikler, som alle er dårlig kvalitet.*

Response: Det er sandt, eller de har uens kvalitet, hvad skal du så gøre med dine patienter? Og det er faktisk dagens emne. På klinikken har vi afprøvet vældig meget, meget erfaring og meget lidt videnskab, og det beror på om jeg er i den situation at jeg læser litteraturen og så tænker jeg - og jeg har været i den situation flere gange - vi trænger ikke til 300 nye eller 600 nye studier, for der kommer disse enorme kvantiteter, og du har jo ikke en chance for at læse det. Så man må jo hele tiden spørge sig selv, "Hvad skal jeg bidrage med, som forsker?"

Question: *Så originalitet?*

Response: Originalitet, plus... de kan være originelle, men de har ikke kvalitet. Det er derfor de får så lav evidens, det er fordi deres forskningsmetodologi, og der kommer vi ind på netop dette, er enten dårligt dokumenteret, du kan ikke læse dig til hvad de har gjort...

Question: *og det har noget med reproducerbarhed at gøre?*

Response: Det har med reproducerbarhed at gøre, og naturligvis statistisk analyse, så det er vanskeligt at tolke, uanset om du laver metaanalyse eller hvad du vil gøre, så bliver det meget vanskeligt. Og for at sige det prospektivt; hvad har vi brug for i forskningskvalitet? Vi har brug for folk, der kan 1) lave systematiske reviews, graduere evidensen og se hvad der ikke er lavet, og siden lave et eller to studier, ikke 600, men de skal laves med en metode som er bedre end hvad der er beskrevet i de her 600. Sådan er det. Altså det er et faktum idag, at enten kan man få lidt metode, eller man bruger forkert metode, eller man har for dårlig kvalitet eller man dokumenterer det sådan, at når man laver disse reviews og vil fremfinde viden om man skal behandle det ene eller det andet, så er det så lav evidensgrad på dette. Så derfor tror jeg det er en rød tråd i forskning, at inden man starter at lave det, og det er en læringsproces, tror jeg, som vi ikke helt har været opmærksomme nok på, eller også har vi ikke sat tid nok af til det. Jeg vil ikke bidrage med et eneste af denne type studier, og havne i den her store bulk med enorme mængder af studiemateriale, som ikke fører til noget. Så forskningskvalitet er for mig, når det kan evidensgraderes som 1. En eller to artikler som er reproducerbare, har metodologi som går op i at forstå, og resultaterne, om de er positive eller negative, det er ikke det. Men det er selve forskningsresultatet som skal være videnskabeligt entydigt. Og der savnes idag den viden og evne til systematisk review og evidensgradering hos mange vejledere. Det tror jeg.

Question: *Hvorfor tror du det er sådan? Man snakker f.eks. om Salami-metoden, at man har eet datasæt, og så publicerer man så mange artikler som man overhovedet kan på baggrund af det. Er det med til at skabe det her problem?*

Response: Ja, det vil jeg tro. Altså, antal artikler er jo et kriterium for at du har et volumen, din forskergruppe har et volumen, og du bliver jo... det er en måde at sige, at det er stor, videnskabelig produktion. Det fremgår sjældent, at man har skrevet artikler med evidensgrad 1. Vi har altså brug for at få et kvalitetskriterie på en eller anden måde. Og det er klart, at citation index siger lidt mere om det her, men om kvalitet, om nogen andre læser det, eller er interesserede. Men vi kommer nok ind på at det er et vilkår, at hvis en forskergruppe engagerer sig i et område, så bliver de ofte citeret, så det siger ingenting om at dette er kvalitativt bedre. Det er nok ikke blevet løst det her.

Question: *Jeg synes det er meget interessant at du snakker så meget om systematiske reviews. Så vidt jeg husker, hvis man ser på evidensskalaen, så er systematiske reviews og meta-analyser toppen. Det er den højeste evidens, og næste trin er randomiserede kliniske forsøg. Stort set alle andre jeg har talt med, de siger at det randomiserede kontrollerede forsøg er det bedste, det er den højeste evidens.*

Response: Det er fordi de fleste beskæftiger sig ikke med systematiske reviews.

Question: *Er det en mangel i uddannelse?*

Response: Det tænker jeg. Det tager meget lang tid inden en forsker idag kommer i kontakt med systematiske reviews. Det må vi jo opdrage fra sundhedsstyrelsen, eller fra videnskabelige organisationer, at man skal være med i det her. Og derfor må man jo uddannes så man kan gøre det. Men har man gjort det een gang kan man bare sige at, uden at gå hjem og gøre alt igen, og lave et nyt systematisk review, og bruge nok så mange uger på at finde al litteratur, det er vigtigt. Det går for hurtigt idag.

Question: *Hvad med at læse dem? Sådant noget som at læse et Cochrane review, er det noget du tror alle kan, eller er det noget, der kræver en særlig øvelse at kunne læse dem?*

Response: Ja på sin vis kræver det lidt øvelse, men hvis man har en journal club i sin forskergruppe, så skal det jo nemmest være automatik, at man går ind i systematiske reviews. Og kommer nogen idag og siger, at "dette kliniske abstract eller randomiserede studie er blevet publiceret, så vi skal gå over til den her behandling", så ville jeg spørge om der findes mere. Men med nye behandlinger findes der jo ofte ingenting, men de kommer jo relativt ofte, og Cochrane findes jo i samtlige lande, disse institutioner og quality insurance, som stort set går gennem tingene. Men som forsker kan man faktisk også tage kontakt til dem og be' om at et systematisk review laves. Og mangler man viden indenfor et emne, evidensbaseret viden, så kan man trigge at der kommer et review. Og det bør være en helt automatisk del af vores forskeruddannelse. For man bliver forbløffet over hvordan forskningsamfundet forvalter sine muligheder.

Question: *Er der forskel på vigtigheden af systematiske reviews, afhængig af om du er kliniker eller forsker? Eller er de lige vigtige for begge grupper?*

Response: De er lige vigtige, det er kæmpevigtigt i dag i klinikken, at man ikke bygger på to randomiserede studier.

Question: *Man snakker meget om, at f.eks. når man skriver kliniske guidelines, så bruger man jo også en evidensgradering, og der er nogen der snakker om, at der er en slags fødekæde, hvor man måske starter med noget basal forskning, følger op med randomiserede forsøg, og på et eller andet tidspunkt skal det gerne ende med en klinisk guideline. I hvert fald hvis systemet virker. Hvordan placerer systematiske reviews sig i den fødekæde? Er de sidestillet med guidelines, eller ligger de inden, eller efter, eller kan det være lidt af hvert?*

Response: jamen det er vel knap muligt at skrive en guideline uden at have alle systematiske reviews i baghovedet. Det ville jeg ikke gøre. Så der findes jo... nej, det kunne ikke lade sig gøre at gå forbi dem.

Question: *Men hvis det nu er et relativt nyt emne, der findes ikke nogle systematiske reviews indenfor det...*

Response: Så må man være så ærlig at sige, der findes de og de randomiserede studier, og man har brug for mere, at man ikke har højere evidensgrad. Og der kunne jeg tænke

mig, at når det er et helt nyt emne, så kommer der mere af det her, "Hvad mangler vi?", så bliver det jo det her, at man med et eller to studier virkelig kan bidrage med noget som savnes. Problemet i dag er, at der er så stor bulk, f.eks. hvis du søger efter bryst- eller prostatakraft, så bliver man forbløffet over at man på 5-6 år har publiceret disse mængder, som er modsigende. Så det er meget mærkeligt. Og hvis man som forsker går hjem og tænker over, hvad er det, der skal til, for at bryde den her trend? Sådan her kan vi ikke fortsætte. Men mange har allerede sat igang med næste protokol, på samme måde. Men det er eftertanke og analyse, og tænke over hvad der skal til, metodologisk, et eller to studier, men ikke hundredevis af denne lave kvalitet. Men det sker fortsat.

Question: *Du har flere gange nu nævnt, man skal tænke over hvad det er, der skal til, hvad det er der mangler. Hvor er der huller i vores viden? Er der publikationer, der fortæller os det? Altså, de her systematiske reviews, er de gode til at pege på, hvad det er vi mangler?*

Response: Altså de er jo ikke en opskrift på, hvad det er du skal gå hjem og gøre, det tror jeg du skal diskutere i forskergruppen. Når man kommer så langt at sige, at nu vil vi gøre det og det, så er det jo at se på, hvad der allerede er gjort. Jeg tror ikke bare man kan gå ud og købe sig en guideline til den type tanke, det er jo også derfor du laver din undersøgelse, vil jeg tro. Vi er opmærksomme på, at kvalitetsbegrebet er vanskeligt. Så skulle du komme frem til nogen guideline for, hvordan man starter bedre protokol, gang på gang, som har originalitet og evidensgraden for sig fra begyndelsen, når man har...[the interviewee starts drawing]... når du lægger et halmstrå på, det er det nye, så må du jo vide hvad det her er. I dag er bulken så stor, og i værste fald er den så heterogen, at du ikke kan sige hvordan du skal bruge det, til en sygdom eller patient, og så inden du bestemmer dig, hvordan ser dit halmstrå ud, som der er ny viden, så må du analysere høstakken. Sådan er det.

Question: *Ja, det var en hel masse om forskningskvalitet helt generelt, og det er superbrugbart. Men jeg vil også gerne høre nogle lidt mere specifikke ting. Først og fremmest relationen imellem forskning og forskningsartikler. Er der en direkte relation, så du kan sige en god artikel altid er baseret på god forskning, men selvfølgelig kan man skrive en dårlig artikel på baggrund af god forskning... men kan man skrive en god artikel på baggrund af dårlig forskning?*

Response: Nej, det tror jeg ikke. Jeg tror det er et tankekors. Der findes jo en hel del forskningsfusk, hvor de skriver gode artikler, og når du forsøger at finde materialet, så er det der ikke, eller metoden er forkert. Og dem har vi jo alle oplevet, og de er kommet frem i lyset efter nogle år. Så jeg tror det ikke. Der findes en del eksempler med... ind imellem er der så brillante artikler, at man næsten bliver lidt tænkensom; "passer det?". Så det tager også lidt tid. Jeg tænker på en artikel, jeg tror det var en japansk forskergruppe, som viste at differentierede celler kan dedifferentieres igen, næsten ned til stamcelle. Og det er jo et koncept som er meget contradictory, så at sige, at man tager en færdigdifferentieret celle og så går tilbage. Så den her forskergruppe, da det

her kom, sad der mange og tænkte, “kan det være sandt?”. Men siden kom der jo reproducerbare studier, og så tænkte man “tak og lov, det er en brilliant artikel, og metoden er reproducerbar”, og dermed er det jo videnskabeligt korrekt.

Question: *Men hvis den metode ikke havde været reproducerbar, så havde det været en dårlig artikel?*

Response: Ja, eller... dårlig artikel... den havde været strålende, men en strålende artikel skal have lidt tid til at sætte sig i sin samtid. Du kan næsten ikke tolke det med det samme. Og derfor ved du også det kan tage 30-40 år inden man får Nobelprisen. Og det er som regel en god beslutning, men der er et lagt delay imellem. Og det skal kunne reproduceres. Jeg kender ingen som har fået Nobelprisen året efter, der skal findes en videnskabelig reproducerbarhed, sådan er det jo i naturvidenskabelig forskning. Så derfor, en god artikel skal modne, hvor den står for sig selv, og siden kommer der to grupper til, så tager man alle tre, og så får de en pris. En stjerne er til tider svær at tolke. Men en god artikel på baggrund af dårlig forskning, den falder.

Question: *Så vil jeg gerne høre lidt om, jeg har nogle begreber, som jeg gerne vil lave lidt brainstorm på, så jeg nævner et begreb, og hvis du så vil sige det første du tænker på. Du må selv om hvor meget du vil sige, om du bare vil sige nogle få ord, eller om du har mere at sige, det er helt op til dig. Det første begreb jeg har, det er peer review. Selvfølgelig i konteksten forskningskvalitet.*

Response: Jeg tænker, at den har sin plads, det tænker jeg. Men jeg ved ikke rigtig... vil du have det på en skala?

Question: *Nej, bare hvad du tænker om det. Er det vigtigt, problematisk, ligegyldigt?*

Response: Nej, det er jo ikke ligegyldigt, det er det ikke. Det er jo en vigtig komponent. Men den ligger ikke helt højt på skalaen, det tror jeg ikke. Men kvalitativt er det helt okay.

Question: *Sådan noget som layout af artikler, er det vigtigt?*

Response: Absolut. Det skal jo følge videnskabelige regler.

Question: *Hvad tænker du på, når jeg siger layout?*

Response: Layout... jeg tænker på et skelet, at være klar og tydelig med strukturen.

Question: *Så introduction, method...?*

Response: Ja. Egentlig tænker jeg på det. Ingen videnskabelig research uden at videnskabelige regler følges. Det tænker jeg.

Question: *Citationer?*

Response: Den er vigtig, index er vigtig. Men citationer, tænker jeg også det er vigtigt, at man angiver sine referencer med tanke på snyd i dag. Det er vigtigt at man ved, det kan granskes idag, det kunne ikke gøres for tyve år siden. Det er vigtigt at man har originalitet, at man ikke anvender andres arbejde uden at citere.

Question: *Og når du giver en reference, er det så også et spørgsmål om at man anerkender tidligere forskning?*

Response: Ja, absolut.

Question: *Impact?*

Response: Ja, impact som sådan, det er det her med research impact, det skal være det her at man nøder sig til at lære sig en gradering af evidens. Hvad betyder det? Og det skal man jo egentlig gøre i sin artikel også, "What does it mean?", discussion og conclusion, men der har man frie hænder idag, og kan måske ikke sige, at det her har en plads eller ingen plads. Men der tror jeg det er bedre, at man gør det her med at sige om der er videnskabelig evidens for noget. Har man ikke mere end grad 3, så kan man sige det, så må det gøres på en anden måde.

Question: *Man skal bare være ærlig?*

Response: Ja, og man må forsøge at lære sig den proces med impact, hvad betyder det?

Question: *Tværfaglighed?*

Response: Vigtigt. Jeg tror man kan kombinere kvalitativ og kvantitativ forskning idag. Og de discipliner som hører til medicin i hvert fald kan i hvert fald være med.

Question: *Det sidste begreb det er ny viden. Det har du vist svaret på.*

Response: Ja, det har jeg vist.

A.2.12. Interview reference: M19

Date: Oct 20, 2011

Time: 08:30

Question: *Du har taget nogle artikler med, dem har jeg bedt dig om at tage med på baggrund af at du vurderede dem som værende høj kvalitet. Har du lyst til at fortælle mig noget om hvorfor?*

Response: Nr. (1) er et review og reviews er rare til overblik, så derfor bruger jeg selvfølgelig en del reviews. For hvis jeg ikke lige kan få overblikket over hvad der foregår i feltet, så benytter jeg det. Den er kort. Og det synes jeg er rigtig dejligt. Så den må ikke blive for lang, selvom det er et review, det kan jeg ikke holde ud at læse, så taber jeg tråden. Den har lidt en drawback i at der ingen figurer er i den, men er så lidt kompenseret af at jeg kender forfatterne og ved at forfatterne skriver meget og i det felt som jeg interesserer mig for og plejer at skrive godt. Og det betyder også noget. Men selvom der ingen figurer er i det her review, og det betyder også noget, så er den dejlig kort og overskuelig og selvom der er meget tekst, så er det stadig overskuelige acceptabel længde afsnit som man kan holde ud at læse. Vigtigst af alt; den er ikke for lang. Det var (1).

Nr. (2) har så nogle af de ting... nej jeg vil lige sige mht. (1), jeg kan rigtig godt lide tidsskriftet, så det betyder også noget det er et tidsskrift som jeg ved publicerer ting der passer til det jeg interesserer mig for, og derfor er det generelt et tidsskrift der popper op tit. Nr. (2) er så måske et tidsskrift jeg ikke benytter så meget, den har så til gengæld noget af det som jeg godt kan lide i andre henseender, den har mange tabeller, og jeg kan godt lide tabeller, det er meget overskueligt og dejligt, og der er figurer, og det er praktisk taget et krav i mange ting, selvom jeg ikke har det til review, så skal der altså være overskuelige figurer i det jeg læser, ellers så gider jeg simpelthen ikke læse det. Hvis det står som tekst, hvis vi snakker kliniske resultater, så gider jeg ikke læse det, så det er såre enkelt. Og så er den her meget kort, overskuelige afsnit, der ikke er særlig lange, og en form som er kendt, sådan at jeg ved hvad jeg kan forvente når jeg åbner artiklen, jeg ved hvilken rækkefølge afsnittene kommer i, så jeg hurtigere kan overskue den. Der kender jeg så ikke forfatterne, men det skal man jo også blive glad og overrasket over en gang imellem, at det ikke altid er nogen man kender. Og så ville jeg faktisk have haft en artikel nr. 3, den skulle nemlig have været en randomized, controlled trial, fordi som kliniker er det det jeg læser det meste af tiden og det er det jeg opfatter som gold standard når jeg skal noget. Men det var sådan lidt svært lige at finde en jeg har læst her for nylig, fordi jeg mest her på det seneste er stødt på nogen som jeg ikke synes er gode.

Question: *Men det vil jeg faktisk også gerne høre lidt om, fordi det snakker jeg jo ellers ikke så meget med folk om - hvad skyldes det så, at du vurderer at de egentlig ikke er så gode?*

Response: Jamen hvis de ikke har fulgt helt den der dobbeltblindede, randomiserede udgave, hvis de f.eks. i sådan en clinical trial er hoppet over hvor gærdet er lavest og har lavet nogle statistiske fiks-fakserier som combined endpoints og sådan nogle ting, som jeg ikke er specielt begejstret for, hvor man putter stroke og alt mulig andet ind i en endpoint, det synes jeg er noget snyd, for det viser ikke det jeg gerne vil have.

Question: *Altså det du mener er hvis der kan være flere forskellige udfald, at man samler det i eet?*

Response: Ja, man samler det i eet punkt, og det gør man rigtig meget i kardiovaskulær

forskning. Så samler man blodpropper i hjernen med blodpropper i hjertet, med perifær karsygdom, og det er snyd. For det er forskellige sygdomme, så jeg synes ikke man kan få noget ud af det. Så de ryger ud til højre igen, for det synes jeg ikke er godt. Så er der også rigtig mange hvor man laver crossover-designs, og jeg ved godt det er fordi så behøver man ikke så mange problemer, men der er tit problemer med washout og alt mulig andet, det synes jeg også er noget hokusfokus, så det ryger også ud. Så jeg må nok sige de der store clinical trials, de skal være store før jeg synes de er gode, så det skal være ordentlige patientmængde, og designet deraf skal være ordentligt. Så kan jeg heller ikke lide hvis de bruger en eller anden vildt avanceret statistisk metode, for så får jeg en mistanke om at de ikke kan hive noget ud af tallene på helt almindelig statistisk metode. Hvis de laver et eller andet vildt fiksfakseri med opdeling i mærkelige grupper, så er det sandsynligvis fordi der ikke var noget da de testede det med helt almindelig gængs statistik. Og det trigger mig også lidt, det synes jeg også er lidt mistænksomt. Så de seneste jeg har stødt på de har ikke været anvendelige.

Question: *Nej, det er jo også interessant at høre hvorfor de så ikke er det. Jeg skal lige høre, det der review, er det et systematisk review? Eller er det bare et ikke-systematisk?*

Response: Nej, det er sådan et lille "hvor er vi nået til indenfor det her interessante felt af forskning".

Question: *En oversigt over litteraturen?*

Response: Ja, for de to som har skrevet det er de to førende på området, så de er sikkert blevet inviteret til lige at give en lille status af hvor vi egentlig er henne med hensyn til det her felt.

Question: *Godt nok, jeg skulle bare lige høre. Hvis vi så skal prøve at se på forskning generelt. Kvaliteten af forskning, hvis vi ser bort fra de her artikler, hvad tænker du der er afgørende for forskningskvalitet?*

Response: For mig er det afgørende når man laver klinisk forskning, at det enten har rigtig mange patienter med, så ens volumen er stort så man kan få fortyndet lidt af de der personmæssige forskelle, og hvis man ikke gør det, så skal man i hvert fald være sikker på at den gruppe man arbejder med er acceptabelt uniform til at man ved hvad man har med at gøre. Det betyder rigtig meget for mig, at der er mange patienter med, når jeg skal synes om det her er noget vi f.eks. skal bruge nede i afdelingen. Og det er ret håbløst, når man er nyremediciner, at få mange patienter. Så det i sig selv sætter en vis begrænsning. Jeg er ret ligeglad med hvorhenne man publicerer, men det siger selvfølgelig sig selv at jeg har visse tidsskrifter jeg ofte hitter på, sådan at når jeg skal finde noget, hvis *Hypertension* f.eks. dukker op meget tit, så bliver det selvfølgelig en slags favoritting, som hvis jeg skal bruge noget i et felt, så søger jeg jo de journals jeg kender. Dernæst går jeg ret meget op i deres design af studier, sådan at designet er ordentligt,

at de kan finde ud af at følge deres in- og eksklusionskriterier og de kan finde ud af at lave deres undersøgelse af patienterne efter bedst mulige standard og sådan nogle ting, så det ikke ligger og svæver lidt i det uvisse hvordan de her patienter de er kommet med og om de er undersøgt ens. Til nogle undersøgelser er det enormt vigtigt at patienterne faster inden, så kan det ikke nytte noget at man vælger at lave dem ikke-fastende og så siger at det godt kan sammenlignes med alle de andre. Så det er ret vigtigt.

Question: *Hvis vi så ser på sådan noget som designtype, altså om det er et randomiseret, kontrolleret, dobbeltblindet forsøg, eller om det er en case...*

Response: ...jeg elsker dobbeltblindet....

Question: *...hvad er det ved det, der er kvaliteten? Er det typen af studiet, eller er det sammenhængen imellem problemstilling og type?*

Response: Sammenhængen imellem problemstilling og type, for der er selvfølgelig andre problemstillinger hvor en randomiseret, dobbeltblindet undersøgelse vil være fuldstændig absurd at benytte. Men til mange kliniske problemstillinger, hvor man skal tage stilling til om den ene pille eller den anden er den bedste til vores patienter er det bare nødt til at være helst en blindet, og om ikke andet en randomiseret undersøgelse. Og jeg synes da klart det er det rareste at kunne, men der er også nogle problemstillinger hvor det enten er etisk fuldstændig håbløst at gøre eller også er det bare ikke sådan problemstillingen er. Nogle gange skal man også bare undersøge patienterne - tage en hel flok og undersøge dem og følge dem og se hvad der sker, og det kan der jo også være værdi i. Men til den kliniske beslutningstagen i dagligdagen da er det lidt de randomiserede der er de bedste, synes jeg.

Question: *Skelner du imellem hvad du opfatter som kvalitet som hhv. kliniker og som forsker?*

Response: Ja, kæmpestor forskel. Fordi det jeg skal bruge nede på afdelingen, der sætter jeg rigtig stor ære i at det er så høj kvalitet, at det er randomiserede trials, hvis der ikke er randomiserede trials synes jeg ikke det er en god idé vi indfører de forskellige behandlinger som folk kommer med løbende, som gode idéer til os. Men som forsker er jeg nok mere eksperimentel, der kan jeg godt gå lidt på kompromis med at det ikke nødvendigvis er sådan i det område jeg arbejder med, det er okay, fordi nogle gange genererer lidt mere speciel forskning jo gode hypoteser til hvad man kan arbejde videre med. Så jeg synes hypotesegenererende forskning kan også være helt fint.

Question: *Men som kliniker kan du ikke rigtig bruge det til noget?*

Response: Nej, det er kun til at nå frem til at der kan laves et trial, som vi kan bruge som klinikere.

Question: *Er det også sådan at grundforskningen er der for forskernes skyld?*

Response: Nej, den skulle helst være der for klinikerne i sidste ende, men jeg kan godt forstå, at der er mange led frem til at der er noget vi andre kan bruge til noget, og en gang imellem dukker der jo også noget op som er interessant at vide, fordi det er interessant at vide. Fordi det ikke skal bruges til noget. Men jeg vil da foretrække, at det grundforskerne finder ud af og opdager og arbejder med, efterhånden som årene går ender med at være noget man kan bruge klinisk. Helst. Men jeg kan godt acceptere, at der er mange led.

Question: *Det var også det jeg tænkte. Der er nogen der snakker om, at der er en fødekæde...*

Response: Jamen det skal der jo være, ellers kan det jo være ligemeget, klinikerne kommer jo ikke fremad hvis ikke grundforskerne finder den ene lille receptor på overfladen af en celle, det skal der til, ellers kommer vi andre ikke videre. Bare jeg ikke skal lave grundforskningen.

Question: *Kliniske guidelines; hvor ligger de henne i det system?*

Response: Jeg hader kliniske guidelines. Jeg synes stadig lidt at kliniske guidelines er lidt sådan nogle "old guys around the table", fordi der er jo, jamen bare tag vores eget hjertestop-instruks, hvis du gennemgår dem, så er der måske en tiendedel af det, som er baseret på reelle studier, og resten er noget vi tror eller deducerer os frem til, og jeg synes stadigvæk at guidelines læner sig alt for meget op ad hvad vi plejer, og ikke så meget op ad hvad der rent faktisk er forsket i. Det vil også være ret svært at lave guidelines, for hovedparten af det vil jo så være noget, der alligevel ikke er undersøgt, som vi bare tror og mener virker. Jeg synes guidelines er noget rod.

Question: *Er det specifikt for det nyremedicinske?*

Response: Jeg synes også det er lidt kardiovaskulært, men det er meget nyremedicinsk, fordi vores patienter er svære og meget anderledes end alle andre, og de er svære at lave trials på, og det er en lille gruppe patienter. Så det der med at lave store, kæmpe trials på en flok nyremedicinske patienter, det er ret svært at komme igennem. Så det er nok lidt sådan i vores felt at vi må forlade os lidt på at sådan er det for nyreraske, så lad os satse på at sådan er det også for dem der er nyresyge. Og f.eks. for ganske nylig gik det jo op for en masse af os, at når man giver kolesterolsænkende behandling til en flok nyremedicinske patienter, så har de altså ikke den samme benefit som en flok nyreraske har ud af den kolesterolsænkende behandling, hvorfor vi pludselig måtte ned og fjerne alle kolesterolsænkende præparater i dialysen, for faktisk viste det, at de døde mere, dem der havde fået det. Hvilket nok var snyd - for vores patienter er så specielle som de er, men alligevel, vi kunne i hvert fald ikke se nogen glæde ved at give dem det. Men vi forlader os meget på andre patientgrupper, og det er noget rod. Så jeg kan ikke lide guidelines.

Question: *Men det er så specifikt for det felt?*

Response: Det tror jeg, og så det der med hjertestop-behandlingen; det har jeg også for nylig været igennem på et kursus hvor vi kiggede på hvad der var belæg for.

Question: *Men er det så fordi de ikke er lavet godt nok?*

Response: Nogle af tingene er bare ikke undersøgt, nogle af tingene er der bare aldrig blevet lavet en undersøgelse af, om det er sådan tingene er. Det handler jo selvfølgelig om at det ikke er særlig nemt at forske i hjertestop, men man kunne jo stadig godt forlade sig lidt på nogle dyrestudier, og der er også for lidt til at vores hjertestop-instruks egentlig er videnskabeligt funderet. Den kører også lidt på hvad vi plejer at gøre, og det plejer at virke... og det er sådan lidt good old guys around the table. Og det synes jeg er helt vildt træls.

Question: *Hvis man skal evaluere forskning, eller kvaliteten af forskning; kan man det?*

Response: Nej.

Question: *Det kan man ikke?*

Response: Det er i hvert fald svært, ikke? For der er mange der har den der med, at den højeste kvalitet forskning er i de tidsskrifter med højest impact, og det passer jo i hvert fald ikke altid. Det er i hvert fald tit sådan at i de helt store, hippe, som *The Lancet* og *The New England Journal of Medicine*, der havner der jo tit noget forskning i, som enten er decideret forkert, eller også som var en fed idé, men når man eftersøger det, ikke er mere. Så jeg tror det er rigtig svært, det er betinget meget af hvor man er henne, og det er betinget rigtig meget af hvilket speciale man har, og hvilken forskningsmæssig interesse, fordi det godt kan være vidt forskellige ting der... nu for eksempel de to artikler jeg har med, det er jo på grund af min forskning. Den der skulle have ligget her som nummer tre, det var den, der skulle have været god til det kliniske. Og den ville være helt anderledes, den gode artikel med sådan et randomiseret clinical trial.

Question: *Tror du der er mere incitament for at snyde sig ind i *The Lancet* end der er for at snyde sig ind i andre tidsskrifter?*

Response: Ja, det kunne da være fedt at komme i *The Lancet*, alle vil jo i *The Lancet*.

Question: *Så det er måske derfor de har problemer med snyd?*

Response: Ja, men det tror jeg faktisk også nogle af de andre har, fordi det jo også kommer an på hvilket felt man snakker om. Hvis man snakker hypertensionsforskning, så er *Hypertension* nok det hippe, og hvis det er arterosklerose, så er det *Circulation* der er det hippe, og jeg tror der er ligeså meget snyd med forsøg på at komme i dem, med

det som de laver, som der er i det andet. Det er også bare fordi *The Lancet* og *The New England Journal of Medicine* er så højranking som de er, at hvis man kunne have en enkelt på sit CV fra *The Lancet*, så ville alle forskere jo være superlykkelige. Og jeg tror det er helt hen i vejret, og det skal i hvert fald ikke vurderes på den måde. Men jeg synes det er svært, også fordi når man laver sådan en pyramide over hvad for nogle designs og studier der er bedst, så står den dobbeltblindede, randomiserede undersøgelse jo bare helt i top, det er the way to go, og der er altså bare ting, der ikke kan undersøges på den måde. Det er rigtig synd for dem som finder spændende associationer, hypotesegenererende forskning, fordi deres forskning er jo mere observationelle studier, og de rangerer altså længere nede. Og det er jo ikke nødvendigvis dårlig forskning eller dårlig kvalitet af den grund.

Question: *Kan de være sværere at få i de store tidsskrifter?*

Response: Uha ja, meget sværere. Det er da meget nemmere hvis man kommer med en randomiseret, dobbelt-blindet undersøgelse. Så kan man da også få nogle trials igennem med meget færre patienter. Det kan jeg da se på nogle af mine medforskere i mit eget felt, som har måske nogle små populationer, hvor de bare lige præcis har fundet ud af, at de vil prøve at give noget medicin eller noget ingenting, og de kan komme afsted med at komme i nogle meget større tidsskrifter.

Question: *Fordi de gør det på den måde?*

Response: Fordi det er en randomiseret undersøgelse, ja. På grund af metoden, og det er jo egentlig snyd, det er jo ikke sikkert deres var det mest interessante for fremtidig forskning, og for at komme fremad. Det kan jo være at det bare var en afrunding af at noget var korrekt, det ved vi så det var sådan det var...

Question: *Det lyder lidt på dig som om du er en lille smule modstander af at man nogle gange så kategorisk siger, at det er det dobbeltblindede, randomiserede, der er guld-standarden?*

Response: Ja, jeg synes den er lækker til den kliniske del, altså nede i klinikken der er det bare det jeg tager frem og siger "fint, den hopper jeg på", men når det kommer til at forske, komme fremad, komme med noget nyt, synes jeg måske ikke nødvendigvis det er det fedeste. Jeg synes ikke der ligger noget nyt, hypotesegenererende forskning i at lave en randomiseret trial, det er jo egentlig bare slutningen, det er sløjfen til sidst der siger at nu virker det. Jeg synes der er meget forskel på hvad der er forskningskvalitet nede i klinikken og når jeg forsker.

Question: *Så har jeg nogle begreber som jeg meget gerne vil høre hvad du tænker om. Sådan lidt brainstorm. Og du må svare lige så meget som du har lyst, eller lige så lidt, positiv, negativ, lige som du har lyst. Selvfølgelig skal det gerne være i konteksten forskningskvalitet.*

Response: Ja.

Question: *Det første begreb er ny viden.*

Response: Dejligt, altid brugbart. Det kan bruges i alle henseender, og alle burde være med til at generere det, fordi det er det, der driver vores hverdag fremad.

Question: *Er det al forskning, der genererer ny viden?*

Response: Ja, næsten, jeg kan næsten ikke forestille mig noget som ikke ville gøre det. Kvalitativ forskning genererer jo også ny viden om hvordan vi skal gøre tingene, og grundforskning, jeg kan ikke forestille mig noget, der ikke genererer ny viden.

Question: *Jeg tænker også på om f.eks. når vi nu snakker kvalitet, hvad så med dårlig forskning?*

Response: Hvis det er metodemæssigt helt håbløst, så er det selvfølgelig lidt den modsatte side. Men stadigvæk, altså, et dårligt opsat studie kan jo også generere noget, det kan f.eks. generere at sådan skal man ikke gøre det for fremtiden, og negative resultater er jo også meget brugbare, og er også ny viden, men det er selvfølgelig rigtigt, at et meget dårligt metodemæssigt studie vil man selvfølgelig ønske var lavet anderledes.

Question: *Hvad med redundant forskning?*

Response: Ah hvad er det?

Question: *Altså hvis nu der er nogen der har lavet et forsøg og har fået nogle resultater, så er det meget almindeligt at man laver et studie eller to for at efterprøve det. Hvad så hvis det er nummer tredivi vi er ude i?*

Response: Fuldstændig spild. Nej, man kan ikke blive ved.

Question: *Men sker det?*

Response: Ja, det tror jeg da. Jeg tror der er masser, der bliver ved med at efterprøve et eller andet resultat, så bliver det måske lidt spild af tid. Det er vel rigtigt, det er måske ikke ny viden. I hvert fald når man når ud til nummer tredivi. Jeg kan måske gå med til nummer fem, og så burde det nok lukkes, hvis man bliver ved med at finde det samme. Så er det nok ikke interessant længere.

Question: *Hvad med sådan noget som layout af artikler?*

Response: Layout af artikler? Hvad er det?

Question: *Jamen det vil jeg jo ikke sige, for det kunne faktisk fortolkes på flere måder. Hvad tænker du på når jeg siger det?*

Response: Jamen layout, der tænker jeg jo straks noget med design, men jeg kunne forestille mig at det måske er en anden slags artikler.

Question: *Nej, det kunne det sagtens være.*

Response: Nå du tænker på hvordan det rent layoutmæssigt er?

Question: *Det kunne det godt, ja.*

Response: Jamen det er da rigtig lækkert at layoutet... altså det er der også ved begge mine, jeg kan godt lide visse ting, når jeg kigger ned over dem, så er det lidt lettere læseligt og så får jeg lidt nemmere budskabet slugt når det er jeg skal. Så jeg synes da det er enormt vigtigt det er layoutet lækkert, men jeg kan da også godt se på den anden side af et lækkert layout af en artikel, hvis der så ikke er noget inde i den. Det er ikke sådan at jeg sluger en artikel, bare fordi den er pæn. Men jeg vil da helst have den pænt sat op. Det gør da min læsning alt andet lige lettere. Tror jeg. Men jeg kunne også godt leve med sådan en artikel som reviewet, der bare er lange søjler dernedaf. Det er altså også okay, bare det der står deri er vigtigt nok og velskrevet. Jeg synes egentlig bare det engelske er mere velskrevet end... jeg vil gerne have det er et ordentlig, velskrevet engelsk.

Question: *Så sproget er vigtigt?*

Response: Ja, det skal være et ordentligt sprog, ellers bliver jeg irriteret, hvis ikke det er skrevet ordenligt.

Question: *Så selv hvis man har skrevet et godt studie, og har nogle gode og vigtige resultater, så kan man måske godt nedjustere den hvis den er dårligt skrevet?*

Response: Jeg kan i hvert fald godt, hvis jeg ikke har rigtig god tid til at læse den og se på den anden side af sproget, så kan jeg egentlig godt komme til at lægge den til side. Egentlig måske uretmæssigt. Der er nødt til at være en slags flow i sproget for at jeg gider læse den til ende.

Question: *Tværfaglighed?*

Response: Ønskværdigt, men måske lidt svært. Men når det kommer til forskning synes jeg det ville være rigtig dejligt. Jeg vil jo faktisk helst at ingeniørerne, radiologerne og alle sådan nogen kommer på banen, så jeg ville ønske at forskningen kunne foregå noget mere på tværs af afdelingerne, på tværs af specialerne og også helt ud. Lige netop det jeg laver, der er det rigtig dejligt at have nogle ingeniører, fordi det er lidt teknisk. Og jeg synes der er for lidt af det, men noget mere af det ville være dejligt.

Question: *Impact?*

Response: Øv. Jeg kigger en lille smule til impact.

Question: *Og du tænker så på impact factor?*

Response: Ja. Fordi det siger selvfølgelig sig selv, at der er nogle tidsskrifter jeg ikke kender særlig godt, og hvis man kigger på impact factoren kan man måske få en lille smule idé om hvorhenne det står, men jeg vil sige, som oftest siger jeg bare nå. Hvis der er en artikel og den er god og den passer til det jeg skal bruge, så er det udemærket at den er publiceret i et low-ranking tidsskrift. Det kan altså ikke røre mig specielt. Og jeg har heller ikke nogen vild idé om at publicere særlig high-ranking. Jeg vil egentlig helst publicere der hvor mit publikum er, og ikke nødvendigvis i et tidsskrift.

Question: *Men er der en sammenhæng imellem de tidsskrifter, der har høj impact og så der hvor de gode artikler er?*

Response: Til dels, jo, for det siger selvfølgelig sig selv at de high-ranking tidsskrifter nok har masser at vælge imellem, så de kan også tillade sig at vælge de bedst skrevne og bedst designede undersøgelser at publicere, hvorimod de andre måske nogle gange må tage til takke med lidt mindre. Men stadigvæk, jeg går ikke nødvendigvis efter impact.

Question: *Hvad med sådan noget som citationer?*

Response: Jeg elsker jo at kigge efter citationer. Jeg synes det er rigtig smart at kunne se en artikels citationer, så jeg kan gå ind i Web of Science og søge fremad, jeg kan se hvem der har brugt artiklen til noget. Dels for at se om det de har brugt den til er noget der passer med det jeg laver, for så kan jeg nemlig sige "okay, så passer den også i mit", og dels så synes jeg også bare det er interessant at se ved en artikel om nogen har gidet at bruge den til noget, eller den har været ligegyldig for alle i fem år efterfølgende. Jeg synes selvcitationer er vildt irriterende, men i nogle felter er det jo nødvendigt, fordi den forskning man laver bygger videre på noget man selv har lavet tidligere, og hvis der kun er to forskergrupper i denne verden der laver det, så kan det ikke rigtig undgås at man kommer til at lave selvcitationer. Og jeg synes i nogle, f.eks. i hypertensionsforskningen, der er det lidt noget rygklapperi. Fordi det ikke er nødvendigt - der er masser af andre man kan citere.

Question: *Peer review?*

Response: Det er en herlig proces... det er rigtig irriterende at skulle igennem reviewers, men jeg synes faktisk sådan en peer-review proces er vældig givende, fordi det er nogle meget kloge mennesker, som har friske øjne på det man laver og kan se nogle ting som man måske ikke selv kunne se da man sad og skrev sin artikel. Så jeg synes egentlig det er okay, nogle af dem er bare snotirriterende, fordi de er meget pedantiske og kører i deres eget område, og vil have man skal citere deres egne artikler, og det synes jeg er lidt noget rod. Men i bund og grund tror jeg at det højner kvaliteten på artiklerne. Det

letter sproget og det højner også det, der står der. Så det er jo godt, jeg har ikke noget imod en peer-review proces. Det kræver bare mange arbejdstimer.

Question: *Siger det noget om kvaliteten, at det er blevet peer-reviewed?*

Response: Ja, det synes jeg. Jeg foretrækker at det er peer-reviewed.

Question: *Det allersidste spørgsmål jeg har det er om du tilfældigvis har været inde og se hvor citerede de to artikler er?*

Response: Ja.

Question: *Før eller efter du valgte dem?*

Response: Efter jeg valgte dem, faktisk. Jeg skulle lige se hvor de lå henne. Den her vidste jeg godt var citeret, reviewet var citeret rigtig mange gange, 60 citationer tror jeg. Også sjovt nok med nogle artikler hvor jeg tænkte, "Det var skægt hvorfor de har citeret den, gad vide hvad det var de havde skrevet deri.", altså der var noget om demens og sådan noget, det synes jeg var sådan lidt specielt. Den her havde ikke helt så mange, jeg tror den havde omkring 20 eller sådan noget...

Question: *Hvornår er den fra?*

Response: Den er fra 2010, okay, det er faktisk også meget pænt. Det er fordi prehypertension er et hot emne, så måske derfor er den blevet benyttet en del. De citerede også den anden artikel, men det fandt jeg jo så først ud af bagefter, da jeg havde valgt den. Det var faktisk fordi jeg havde valgt den til noget arbejde, at den dukkede op. Laurent & Boutouyries område kan ikke indgå at de selv citerer meget, de er koryfæerne indenfor området, så de har også citeret til den her selv efterfølgende, så sådan må det være. Og det gør jeg altid, på alle mine artikler.

Question: *Jamen det var sådan set det jeg havde så, medmindre der er et eller andet du vil tilføje.*

Response: Næh.

A.2.13. Interview reference: M21

Date: Oct 26, 2011

Time: 10:30

Question: *Hvis du vil starte med at fortælle mig en lille smule om hvorfor du har taget de her artikler med?*

Response: Ja, den ene (1) er en af de bedste artikler jeg har mødt i min tid. Det er et studie som er lavet meget, meget grundigt. Det er et dobbeltblindt, randomiseret studie hvor man har prøvet at behandle Hepatitis B patienter. Man har så på forhånd besluttet nogle bestemte endpoints, som man ville holde øje med, man har så yderligere valgt, at der sidder en komité udenfor studiegruppen, altså nogen der ikke er involveret i studiet, og evalueret de endpoints og stopper, når man siger, at nu er der tilstrækkelig med endpoints. Det er et ekstremt gedigent studie; det er faktisk et af de eneste studier der har vist, i den her sammenhæng, at det kunne svare sig at behandle, måske endda det eneste, og det kan endda være at der ikke må laves flere studier herefter, fordi det er så grundigt som det er. Det er et placebo-kontrolleret studie og det er det der gør det specielt i den her sammenhæng, for det kan man ikke få lov at gøre længere, for nu har man påvist, at der er en behandling der virker. Og derefter er det den, der er minimum i hvert fald. På alle måder et velgennemført studie. Det er efter bogen.

Question: *Så man har simpelthen taget flere og flere patienter ind, indtil der var nogen der sagde "nu har vi nok"?*

Response: Nej, det er den anden vej, man har taget ca. 600 patienter, jeg kan ikke huske om det er 6- eller 700 patienter, med meget svær Hepatitis B, som det er det her handler om, så har man givet den ene halvdel placebo, og den anden halvdel det aktive stof. Så har man fulgt de to grupper, og så har man hver gang, der har været et event, det kunne være de fik leverkræft, det kunne være de fik vand i bugen eller hvad det nu er for nogle ting man får, eller blødende åreknuder i spiserøret, som er nogle terminale ting i forhold til det her, og så har man så talt op, og på et eller andet tidspunkt - og det var faktisk, det var meningen det her skulle forløbe i fem år, og så vidt jeg husker var det efter ca. 2 1/2-3 år man stoppede studiet, fordi der var der så stor forskel på de hard end points som man havde valgt at gøre op på. Så det er et helt utrolig godt gennemført studie. Det er sådan noget, som Cochrane ville drømme om.

Question: *Godt... hvad med den anden?*

Response: Den anden har jeg taget med, fordi jeg synes den er interessant. Det er et helt nyt område, man er begyndt med at se på, noget som man egentlig har kunnet i mange år, men som ikke har kunnet se mulighederne i, eller brugen af, og der er det så en tysk gruppe, som også igen - det er ikke et behandlingsstudie, det er noget med at se hvordan man kan bruge den markør som man nu har taget op, og der har de så gået igennem og appliceret den på alle de tænkelige situationer hvor den kunne være brugbar, og også igen gjort et meget grundigt arbejde, altså taget alle de ting, som jeg er interesseret i, i det her område, kunne ønske mig at de skulle undersøge, og det har de gjort og samlet sammen på en meget grundig, gedigen måde. Jeg kunne have valgt mange andre, men det var ligesom to som... den her (1), det vidste jeg med det samme, da du skrev, at det skulle være det ene, for det er bare et knaldgodt studie, og jeg bruger det også når jeg underviser på ph.d.-kurser, for at vise hvordan man kan gøre, til forskel

fra hvordan det ofte bliver gjort. Så derfor... det andet er mest fordi jeg synes det er interessant, og de har gjort det grundigt, og de har gjort det rigtige, de har været ude i alle kroge og taget det hele med, og været ude i hver eneste situation og gjort op, "Kan man bruge det her?", og så er de nået frem til, at det kan måske bruges. Meget reelt. Hvor der var en fransk gruppe, der lancerede det her og buldrede frem, det var verdens 8. vidunder, og nu skulle vi til at gøre det allesammen, og det må man sige, det var det nok ikke, men det kan nok bruges til nogle ting - og det har tyskerne så vist, at i de situationer kan man bruge det, og i andre situationer er det helt meningsløst. Det var ligesom mine bevæggrunde.

Question: *Det handler meget om design af studiet...*

Response: Ja, altså, man kan sige det her (2) er måske ikke så meget design - det her, der har man valgt at tænke sig rigtig grundigt om, inden man gik igang, og har taget alle de patientgrupper med, man kunne have interesse i at få belyst. Det andet (1) er helt klart design - alting er på plads, styrkeberegning er på plads, alt er i orden, de har gjort alt det man skal. Og så har de også... etikken har de overholdt, spurgt alle patienterne, så alt er bare i orden der, efter min mening. Så derfor er det sådan et skoleeksempel på god forskning.

Question: *Er der andre ting ved dem, andre karakteristika, som du vil sige, det er også et tegn på kvalitet?*

Response: Man kan sige det her studie (1), det at man har valgt at tage en gruppe udenfor de aktive forskere, udenfor de engagerede personer til at vurdere outcome, det er efter min mening helt fantastisk godt lavet. Fordi det er jo lige præcis der det svage punkt er, ofte, fordi forskerne vil jo gerne, måske ubevidst, vise at behandlingen virker, eller den nye behandling virker, eller hvad nu... så der er jo en bias der ligger indbygget i at man er engageret i det man laver. Så derfor er det en utrolig god ting, at have nogen som sidder helt udenfor, som i princippet er helt ligeglade med, hvad det her studie viser. Om det viser det ene eller andet er de kolde i forhold til. Og de er så faktisk blindet i forhold til hvem de her resultater kommer fra. De ved ikke, om det er en placebo, eller en medicinbehandlet patient, de sidder bare og tager stilling til, hvilken gruppe det er, der bon'er ud, og hvornår er det man skal sige stop. Det eneste, hvis man skal sige ét lille kritikpunkt, det var at de havde valgt en række end points, altså de havde valgt leverkræft dekomensation, dvs. leversvigt, så falder den af, så et andet punkt er blødende årer i spiserøret, eller det her med vand i bugen, og jeg kan ikke huske om der var et mere, død, selvfølgelig, og det man egentlig gerne ville vide noget om, det var leverkræft, kunne denne behandling forebygge leverkræft, og lige netop der er styrken måske... den er statistisk signifikant, men det er den svageste af styrkerne i det her, det er de andre, der er bon'et ud først, så derfor er studiet blevet stoppet på et tidspunkt, hvor jeg måske godt kunne have tænkt mig, hvis nu det havde kørt et år mere, så kunne man have set, kunne det faktisk forebygge canceren på en anden måde?

Det er det eneste kritikpunkt, men selv der er det stadig statistisk signifikant, det er ikke det, det er bare den svageste signifikans i forhold til de andre.

Question: *Du nævner etik, er det noget... hvordan er det egentlig i sundhedsvidenskabelige forskningsartikler, er det noget man nævner? Tit?*

Response: Ja, og mere og mere. Man kan sige, i dag må der jo ikke foregå et behandlingsstudie uden der er informeret samtykke, informed consent, og det skal fremgå, og det skal også fremgå, at patienterne har fået en skriftlig og en mundtlig orientering om, hvad det er de deltager i. Og jeg kan sige, jeg er jo så sommetider reviewer på andres artikler, og der havde jeg en for nylig, hvor jeg måtte sige dels - det var et meget stort studie, hvor de havde 6,000 patienter over en årrække, og der var både børn og voksne, og hvor de anførte, at alle var spurgt - og der må man sige, sådan som de havde bygget det op, der var det usandsynligt, at man skulle kunne have spurgt alle de her, i hvert fald givet skriftlig information og fået et skriftligt samtykke. Så det måtte jeg som reviewer sige til editor, at jeg anså det for at være tvivlsomt om det var sandt. Og det er jo noget man tager meget alvorligt i vore dage, til forskel fra for 10-20 år siden, der var det noget andet. I dag er man meget fokuseret på det her, og det synes jeg er meget reelt, fordi der har været meget kritik af, at man lavede studier i tredjeverdenslande, og de ikke vidste hvad de blev udsat for, og meget af det afprøvning der var af HIV-medicin i gamle dage i Afrika, hvor de så bare fik noget, og så stoppede man igen, det var jo vældig kritiseret, så det synes jeg er super.

Question: *Er der en sammenhæng med kvalitet, eller er det bare to forskellige ting?*

Response: Jeg mener da helt klart, der er en sammenhæng med kvalitet, for hvis man har tænkt det hele igennem, og også tænkt etikken igennem, så får man også et produkt der er bedre. Det er helt sikkert, der er helt klart en sammenhæng imellem dem. Det skal være en del af det, for det gør jo, at man også tænker det meget grundigere igennem. I Danmark er der jo en række krav, og det er der også - jeg er med i en gruppe der laver studier i Afrika - der har de jo også deres etiske komitéer, og der er meget kritiske i forhold til om der kommer nogen udefra og laver noget, om det er godt eller skidt, om etikken er i orden, fordi det skal jo nødvendigvis være sådan at man bagefter kan klandres for at det er sådan noget kolonimagtsforskning eller sådan noget.

Question: *Hvis vi så i stedet for de her to artikler ser på forskning i det hele taget, helt generelt forskning, og jeg så siger "forskningskvalitet", hvad tænker du så på?*

Response: Jamen jeg tænker på forskellige ting. Jeg tænker selvfølgelig på at vi har sådan et gradueringsystem, med hvad der har den tungeste vægt. Og der er sådan noget som Cochrane reviews eller meta-analyser sådan nogen som der ligesom ligger øverst på hitlisten.

Question: *Og det er evidensniveauer?*

Response: Det er evidensniveauer vi snakker om, ja, og selv der skal man jo være kritisk. Og igen må jeg sige, at det er jo ikke alt, fordi der nu står Cochrane på det, eller fordi der står meta-analyse på det, at det bliver verdens fineste videnskab, det er det bare ikke. For hvis det er noget skidt man putter i det, så er det også noget skidt der kommer ud. Så det er meget afhængigt af de forskere og den kritiske linie de selv anlægger. Så er det klart, at hvis vi snakker behandlingsstudier, så er det sådan nogle dobbeltblindede studier, randomiserede studier som er de vigtigste, og som også er dem, der flytter noget. Men igen skal de jo være designede til det, de skal vise. Man skal gerne have formuleret sine forskningsspørgsmål så grundigt på forhånd, så man også er sikker på, man kan svare på det. Og det er jo tit det er udført med een tanke og så er det noget andet man finder ud af, og så bliver fokus ligesom drejet over på, at det også var spændende. Sådan noget skal man også være meget kritisk i forhold til; hvad lagde de ud med, og hvad endte de med.

Question: *Det er faktisk noget jeg hører rigtig mange sige, det her med de bedste studier er dem, der er blevet designet ordentlig fra starten og hvor de resultater man ender med, det er også dem man egentlig satte sig ud for at undersøge. Men de siger også, ligesom dig, at der er også nogen der gør noget andet, de finder noget hovsa-agtigt undervejs, eller de laver data-mining på nogle datasæt. Er der noget vi kan gøre for at øge kvaliteten, for at sikre, at...*

Response: Altså, det er vel ikke... de der hovsa-resultater, hvis vi kan kalde dem sådan, de er også interessante, det skal bare klart fremgå, at det var noget andet end det vi lige gik efter, men vi fandt altså det her, det skal nok undersøges i en anden sammenhæng. Jeg tænker at i den her sammenhæng er der vel ikke andet for end uddannelse; at folk ved hvad det er de beskæftiger sig med. Og der tænker jeg i hvert fald at en forsker, ikke nødvendigvis en forskeruddannelse, men en uddannelse i at læse en artikel kritisk, det er utrolig vigtigt. Det der med at man ikke bare læser de første tre linier og så tænker, "Nå, er det sådan?", eller får den stukket i hånden af en der siger hvad det er, og man så tænker, "Nå, nu ved jeg også det.". Man bliver altså nødt til at sætte sig grundigt ind i præmisserne for at vide hvad det handler om. Og selvfølgelig - jeg læser et hav af artikler som jeg bare lige skimmer, men hver gang du skal bruge det til noget fornuftigt, så skal du jo sætte dig ned og se på, "Hvad er det de har lavet, hvad har de tænkt, hvad har de gjort, hvad har de fået ud af det, og hvad konkluderer de så på det?". F.eks. er der nogle ting, altså indenfor mit felt indenfor klinisk biokemi, der er der sådan nogle markører, som måske i tyve år har været forsøgt anvendt til noget, og der er ligesom ikke kommet gennembrudet. Så kan man sige, der er kommet 80 artikler, firmaet har sponsoreret de 70, det er den samme sidsteforfatter, eller det er den samme forfatter der går igen på over 90% af dem, sådan nogle ting, hvis man ikke sætter sig ind i det, så bliver man snydt, så tror man det er skidegodt det de har lavet, men så ser man at han også får løn af firmaet, og sådan noget, og så tænker man, at man nok lige skal tage det med et gran salt, og kunne der ikke være en uafhængig gruppe, der havde lavet noget på det her. Så det er sådan nogle overvejelser, det tror jeg kun man kan lære, hvis man får forståelsen

for hvor vigtigt det er, at læse kritisk. Og altså, læger lærer det jo efterhånden. Det var ikke noget vi lærte da jeg gik i skole.

Question: *Men for lige at gå tilbage til det med evidens, så siger du, at man ikke bare ukritisk kan bruge den skala, uanset om det er Cochrane eller meta-analyser, så skal man stadig være kritisk. Men er der en generel sammenhæng? Kan man sige, lidt forenklet, jo højere man ligger på evidensskalaen, jo bedre kvalitet signalerer det?*

Response: Man kan sige, de to øverste niveauer... i virkeligheden vil jeg sige at den kontrollerede, blindede undersøgelse, det er den jeg sætter højest. Det er klart, hvis der er en række studier som ligner hinanden, som man kan sætte sammen, så man kan bruge data fra dem, så kan man gøre det. Men jeg har også oplevet at man i Cochrane-sammenhæng drog nogle konklusioner på noget, som var abstracts som måske ikke havde været i stand til at få publiceret på syv år, altså, på en eller anden interessant problemstilling havde de på trods af at de havde lavet et studie i 2002 ikke fået det ud i 2009. Så mener jeg også man med sin kritiske sans bør sige, "Hvad er det for en type undersøgelse, som ingen tidsskrift vil tage, på syv år. Så der er sådan nogle ting... og hvis man putter det hele sammen og siger vi har 14 abstracts og 15 klinisk kontrollerede undersøgelser, og så lægger vi det hele sammen i en pællevælle, og så siger vi, at vi kan drage en konklusion på det - det mener jeg ikke man kan. Det mener jeg ikke er ordentligt. Men det gør Cochrane rent faktisk - så der skal man være rimelig kritisk i forhold til det. Men derimod kan man sige, at hvis du har en patientgruppe som er undersøgt med ti forskellige studier, som i virkeligheden ligner hinanden rigtig meget, og hvis man nu tog alle patienterne ud og lagde dem i en database, og så kiggede på dem samlet, så kunne man tillade sig at konkludere noget, det synes jeg er okay. Og det giver en høj grad af sikkerhed.

Question: *Relationen imellem artikler og forskning, hvordan ser du den? Er det sådan at artiklerne også afspejler kvaliteten af studiet, eller er der stor diskrepans der?*

Response: Det skal de jo gerne, ellers er der noget der er gået galt... hvis man har lavet et skidegodt studie, og artiklen den bliver dårlig, så er der jo noget, der er kikset. Og på den anden side er det nok svært at lave en skidegod artikel hvis man har lavet et dårligt studie. Så jeg vil sige det hænger på en eller anden måde sammen. Så jeg vil sige, hvis man laver gode studier, så laver man også gode artikler.

Question: *Så det med at evaluere kvaliteten af forskning på baggrund af artiklerne er ikke helt hen i vejret?*

Response: Det synes jeg ikke. Det synes jeg er helt okay. Men nogle gange vil du også opleve, at hvis du læser en artikel, at de f.eks. ikke har beskrevet den metode de har brugt. Og så kan man sige; hvis man sidder og kigger på nogle resultater, og man ikke ved hvad metoden er, hvad er det så man kan konkludere på det? Og hvad kan det så

skyldes? Har man fået for lidt information, eller har man nogle gange bevidst fortiet det man har... måske også fordi tiden er løbet fra... hvis man har lavet det her studie for fem år siden og skal have det publiceret så i mellemtiden er metodologien revolutioneret, så er det rigtig svært. Og det er rigtig svært at skrive, at man har brugt den næstbedste metode til det her studie. Så der er sådan nogle overvejelser, og det skal man selvfølgelig også kigge efter, men ellers vil jeg sige, normalt, hvis man giver alle de oplysninger der er, og det er så ikke sikkert editor vil have alle oplysninger med i artiklen, men hvis det så er i et bilag bag i artiklen, eller man på nettet kan finde ekstramaterialer, så er man rigtig godt med.

Question: *Og det er blandt andet sådan noget som reproducerbarhed det går på?*

Response: Ja, reproducerbarhed er jo et af de meget vigtige emner. Men det synes jeg i det hele taget at reviewer og editor bør sikre sig, at resultaterne er reproducerbare, for det er jo ikke interessant at vide, at man har kunnet finde et eller andet, hvis man ikke kan genfinde det. Det er faktisk ikke rimeligt at publicere sådan noget. Det er sådan noget man kan stå og holde et foredrag om, og sige, "Nu skal vi hjem og se om vi kan gøre det een gang til.", det er okay, men det er ikke okay at publicere det. Jeg synes man skal have data for sin reproducerbarhed, også i den sammenhæng, at metoden dur.

Question: *Er det et problem i dag? Er der mange forskningsartikler, hvor det ikke er reproducerbart, enten fordi resultaterne ikke ER reproducerbare, eller fordi metoden ikke er beskrevet godt nok?*

Response: Man kan sige det kommer an på hvad det er for en type studier du kigger på. Hvis du kigger på de store behandlingsstudier som der jo er mange af nu, så mener jeg ikke det er et problem. Det er ikke der problemet er. Men hvis du kigger på metodeudviklingsstudier, så kan det være et kæmpestort problem. jeg arbejder jo med sådan noget hvor man også udvikler metoder, og hvis vi nu kigger på et eller andet fra f.eks. Japan, så har vi sådan en holdning til det, at der er et eller andet, der mangler. Der er altid et eller andet de ikke siger, fordi det også er sådan noget med patenter, så der er altid en lille hemmelighed et eller andet sted. Og det er jo lidt et holdningsspørgsmål, om det at man publicerer er en ægte videndeling, at man gerne vil det, at andre skal kunne det, man selv kan. Eller om man faktisk prøver at forholde andre muligheden for at gøre det samme. Det er to forskellige problemstillinger. Og der mener jeg, jeg hører til den skole der mener, man bør fortælle det man gør, og så nøjagtigt, at andre kan gå ud og reproducere ens data. Men det er ingeniørens situation i dag. Men det er også et holdningsspørgsmål. Og nu er vi jo så ude i alt det her med at man skal tænke i patenter og alt sådan noget, og det er jo et eller andet i modstrid med den fri videndeling som kunne være ønskelig, ik? Men det er jo også noget der kommer... hvis man får bevillinger fra forskningsrådet, så er der jo også krav om at man skal lægge sine data ud, andre skal have lov til at bruge det. Så det er jo noget, der kommer mere og mere og som er åbent. Det ønsker man også.

Question: *Ved du tilfældigvis hvor ofte de her to artikler er citeret?*

Response: Nej... den her (2) tror jeg ikke er citeret ret meget. Den her (1) tror jeg er citeret helt vildt meget. Det er måske en af de mest citerede artikler på hele Hepatitis B området, tror jeg. Den anden (2) er mere for feinschmeckere, det er for nørder, så det er ikke sikkert den er citeret ret mange gange. Det tror jeg faktisk ikke.

Question: *Det sidste jeg har er en kort brainstorm, hvor jeg vil nævne nogle begreber, og så vil jeg gerne have at du svarer med din umiddelbare reaktion på det. Kort eller lang, det bestemmer du selv. Og det er selvfølgelig i konteksten forskningskvalitet. Det første begreb jeg har det er ny viden.*

Response: Ja, altså, det vi går efter i al forskning det er jo ny viden. Det er ønsket. Det lykkes også en gang imellem.

Question: *Tværfaglighed?*

Response: Altså jeg går ind for at man er meget tværfaglig, at man bruger mange forskellige indfaldsvinkler til at se på problemstillingen. Og jeg prøver også i min egen forskningsgruppe at få forskellige personer ind, som har forskellige synsvinkler, for det skaber altså en flerhed som er meget større end hvis man er syv læger, der sidder og snakker om det samme. Så der mener jeg også at tværfaglighed i den sammenhæng er sygeplejeforskning og sådan noget forskning som du er igang med, og alle mulige andre som er igang omkring komplekset. Det synes jeg er utrolig væsentligt.

Question: *Hvad med den anden vej rundt... hvordan skal man sige det... det at forskning indenfor et felt kan bruges indenfor andre felter... skiller det sig ud?*

Response: Tænker du metode eller konkret indhold?

Question: *Jamen begge dele, sådan set.*

Response: Man kan sige, ved metode der er det jo interessant, for vi har jo forskellige metoder. Hvis du tænker på hvordan de laver forskning på Humaniora, så vil der nogle gange være nogle meget bløde forskningsmodeller, også med sådan noget som du er igang med her, sådan en interviewmodel, det er jo sådan noget som er ganske anderledes end den hardcore naturfaglige skole jeg kommer fra. Selvfølgelig kan man bruge begge dele. Noget jeg synes jeg har oplevet er, at man også godt kan anlægge den der metodologiske anskuelse som man bruger i det naturfaglige, den kan man godt nogle gange bruge med stor fordel også ovre i den mere bløde forskning. Og dermed, efter min mening, kvalificere den lidt mere end ved at man kun har fået bløde input. Og det synes jeg er noget af det man godt kunne lære. Min kone har skrevet nogle ting i forbindelse med noget ude på Aalborg Universitet, noget læringshalløj, og der har vi så diskuteret netop om man så kunne anlægge den her metode. Og det strammer jo

enormt meget op på sådan nogle studier hvis man gør det, fordi opskriften er så stram. Men man kan også nogle gange med fordel gøre det, ikke at man kan gøre det altid, men man kan godt bruge forskellige metoder på forskellige områder. Og jeg synes også man skal være fræk og gøre det og føle sig fri til at gøre det, men ikke sådan så det skal være den videnskabelige metode vi bruger, som skal bruges på alt, det kan også være den anden vej rundt, så man kan bruge de andre metoder til også at bedrive forskning indenfor den lægevidenskabelige eller naturvidenskabelige. Det var den ene halvdel af dit spørgsmål, hvad var den anden halvdel?

Question: *Det var mere på resultaterne.*

Response: Ja, det er helt klart at hvis jeg... det kan være svært at drage nogle faste konklusioner, nogle faste konklusioner, på sådan en bred... hvis du nu har lavet et fokusgruppeinterview med otte personer, og så har de snakket om nogle bestemte emner, som de har fået besked på, det var det de skulle snakke om, og så drager du nogle konklusioner. Det kan være enormt svært, for det kan ligeså godt være otte forskellige konklusioner der kommer ud af det. Og derfor tænker jeg det er meget vanskeligere end f.eks. to punchlines i den form for forskning som jeg så bedriver.

Question: *Det næste jeg har er impact.*

Response: Jamen alle vil jo gerne have at deres forskning giver impact. Men på en måde synes jeg det er for opreklameret, at man hele tiden skal gå efter højeste impact. Nogle gange vil det være - hvis man nu laver et gedigent, dansk studie, så vil det være mest egnet for danskere, skandinaver måske, nordeuropæere kunne man sige, man kunne generalisere ud til. Og det vil jo ikke give en kæmpehøj impact. Men det har jo i den sammenhæng hvor det er tænkt en fantastisk impact, men du kan ikke måle det på den store skala. Så derfor tænker jeg sommetider, at det der med at man altid skal gå til, og jeg gør det jo selv, det er jo sådan noget vi er præget af, det der med at man går efter højeste impact factor, og så rykker man nedad i rækkerne. Men i virkeligheden kan jeg jo godt li' at publicere i skandinaviske tidsskrifter, fordi det giver en anden tilfredsstillelse, at data bliver anerkendt af dem, man egentlig samarbejder med, og som man gerne vil have skal vide det, og et eller andet sted er det måske lidt ligegyldigt, om der sidder en Amerikaner et sted og har læst det eller ej - det har jeg jo aldrig om alligevel. Så det er sådan...

Question: *og så bruger du jo faktisk ordet impact på to forskellige måder nu. Du siger både impact factor, men det lyder også som om du siger, at det egentlig har en betydning.*

Response: En reel impact, ja, og det opfatter jeg også som to forskellige ting. Vi bliver også vurderet efter impact factor, men i virkeligheden kan det godt være vores studie har kæmpe stort impact fordi det indenfor det område det drejer sig om er meget relevant. Altså, hvis man nu kigger på tarmsygdom blandt skandinaver, så kan det være det

er fuldstændig anderledes fra hvad det er i Afrika eller Mellemøsten, derfor kan man sige, det vil aldrig få impact, altså påvirkning, af deres måde at se det, men det kan få stor påvirkning i vores område. Og der kan være nogle felter - det jeg arbejder med er jo mest Hepatitis, der adskiller Amerikanerne sig meget fra mange af os andre, i hvert fald nordeuropæerne, og derfor så er det måske heller ikke så interessant for dem, og så opnår man heller ikke den højeste impact i betydningen impact factor, fordi du ikke får publiceret i det amerikanske tidsskrift, men du kan opnå glimrende impact i virkeligheden, der hvor du er, det er bare ikke målbart.

Question: *Hvad med sådan noget som citationer?*

Response: Det må jeg sige, det interesserer mig virkelig meget, meget lidt. Det er klart, det ville jo være det ideelle, om ens artikler også blev citeret. Men det er ikke et område, som interesserer mig ret meget.

Question: *Du har ikke nogen fornemmelse af, hvorvidt der er en sammenhæng imellem de gode artikler, og antal citationer?*

Response: Altså jeg har en fornemmelse af, at citationer kan være bedrageriske. Der var jo et af de studier, som ligger til grund for, der er mange studier i virkeligheden, som har været voldsomt citeret, som viser sig at være scientific fraud. Så derfor skal man lige tænke sig om i hvert fald. Og et af de studier omkring frugt og grønt, og som var lavet i Indien, og som bruges af alle til at sige det er sundt at spise frugt og grønt, hvor det viste sig ti år senere, at det var fup og svindel. Og så kan man så sige, at det er noget man skal tænke igennem. Så det er ikke... det lægger jeg ikke så meget i.

Question: *Okay. Så har jeg et begreb der hedder "layout af artikler"?*

Response: Ja. Det er jo nærmest standard hvordan de skal være, og det overholder vi jo til punkt og prikke. Det er jo de tidsskrifter vi sender til, der bestemmer hvordan de skal se ud, og vi indretter jo vores artikler efter dem. Og så en gang imellem, det sker jo næsten hver gang, når man så får artiklen tilbage fra det første tidsskrift, så skal den skrives om i forhold til det andet, og måske med en lille smule anden vinkel, for at den passer ind i den linie de så har. Og det er jo noget vi så tilpasser, og modellen er jo stort set den samme, men vi tilpasser i vid udstrækning, for at det skal ligne det de vil have... det vi tror de vil have. Og du ved, efterhånden på Internettet, når du lægger det ind, det er jo fuldstændig sat op hvordan det skal være, så du har jo ikke mange frihedsgrader. Der er jo også hvor mange ord og hvordan og hvorledes, så alt i alt er der jo en stor styring udefra i forhold til det du selv kunne tænke dig at fortælle i dit studie. Du bliver også nødt til at barbære noget væk.

Question: *Det sidste begreb jeg har, det er peer review.*

Response: Yes. Peer review, det er jo noget som vi bruger meget tid på. Nåde at

efterkomme peer reviewere, men også at være det selv. Jeg synes det er en udmærket ting. Jeg synes man skal være meget kritisk med sin egen måde at gøre det på, og netop hele tiden tænke over, om man er kritisk fordi man i virkeligheden heller ville have publiceret det her selv, for man får jo altid noget, som man er ekspert indenfor, det er jo ligesom altid ens egne områder, eller om man er kritisk... man skal i hvert fald være meget bevidst om hvorfor man er kritisk. Og jeg synes man skal være konstruktiv kritisk, hvis det er muligt. Nogle gange kan man også sige, at det simpelthen er for dårligt, det har de ikke gjort færdigt inden de sendte det ind. Og så kan man sige, det vil jeg ikke bruge min tid på. Men det er sjældent. Jeg synes langt de fleste gange kan man få nogle meget konstruktive review-kommentarer, og også nogen som kan være med til at gøre ens studie bedre. Altså, der sidder nogen, som aldrig har set det før, læser det, og så tænker nogle gode tanker og det kan det blive meget skarpere af. Altså, jeg har lige for nylig publiceret en artikel, hvor jeg havde en reviewer, jeg tror det var en jeg selv havde anbefalet, og som så er en af verdens førende på det her, han kom med nogle knaldgode kommentarer, han gjorde artiklen betydelig bedre, ved at vi efterkom hans kommentarer og gjorde det han sagde vi skulle. Så det er jo ligesom at have en mentor, kan man sige, det er bare en anonym mentor i den her sammenhæng.

A.2.14. Interview reference: M25

Date: Oct 28, 2011

Time: 15:00

Question: *Du har taget tre artikler med, fordi du har vurderet dem som høj kvalitet. Vil du fortælle mig lidt om hvad det er ved de tre artikler, som du finder af høj kvalitet?*

Response: De er taget med ikke kun fordi de er af høj kvalitet, men også fordi jeg mener de eksemplificerer noget principielt. Nr. (1) det er en kasuistik. Det vil de fleste jo mene, det er røven af fjerde division, og ikke anvendeligt til noget som helst. Men den illustrerer faktisk at kasuistikker også kan bruges til noget. Det er en helt ekseptionel historie, hvor en vordende læge oplever at han får kvalme og noget der kunne ligne leverbetændelse, den gang han er på job som anæstesiolog. og så i hele hans erhvervskarriere, i hvert fald som yngre læge, der går han til og fra det speciale, og dermed til og fra udsættelse for Haluthan. Og så er han så så heldig, at han arbejder sammen med en leverlæge, som on and off i hans eksponeringer tager leverbiopsier og blodprøver, så rent faktisk kan den ene person og hans forløb over adskillige år - der kan han helt og aldeles ud fra en kasuistisk meddelelse fastlægge hvad var og er haluthant hepatitis. Det er et eksempel på at man ikke altid skal sige at en kasuistik kan man ikke bruge til noget. Så det er begrundelsen for at jeg har taget den.

Question: *Og den er jo faktisk fra New England Journal of Medicine.*

Response: Ja ja, syv sider i NEJM, og så er eksponering og forløb og det hele i sådan

en sygdom fastlagt.

Question: *Hvor almindeligt er det, at NEJM bringer så lange kasuistikker, eller i det hele taget kasuistikker?*

Response: Aner det ikke. Jeg kunne godt forestille mig, at det aldrig er sket før. Det vil jeg egentlig gætte på. Og jeg bruger den jo i undervisningen, det er derfor jeg lige kan hive den frem fra ringbindet. Jeg har brugt den i undervisning, rent metodemæssigt, som et eksempel, udover thaladomid-sagen, om kasuistikker som “the unexpected, where discoveries begin”. Som modvægt til de folk, der postulerer at der ikke kan siges noget som helst fornuftigt, hvis det ikke er i et randomiseret forsøg med kontrolgrupper og alt det der. Og det er selvfølgelig ikke det samme som, at enhver kasuistik er banebrydende, men jeg mener den her er en helt ekseptionel, god og væsentlig artikel. Antallet af kasuistikker i NEJM er formentlig meget tæt på 0.

Question: *Hvad med de to andre?*

Response: Så er der jo en klassiker. Det er så en case-control undersøgelse; Richard Dolls omkring tobak. Den mener jeg også er interessant af flere forskellige grunde. Først metodemæssigt, for det er vel en af de første case-control studier, som blev lavet på et tidspunkt hvor begrebet ikke var opfundet. Det illustrerer dernæst, det fandt forfatteren selv ud af, mig bekendt, at han egentlig havde lavet et forkert metodevalg, af kontrolgruppe. Det offentliggjorde han så med en anden kontrolgruppe et års tid eller to senere. Og derudover er det startskuddet, i hvert fald vurderes det sådan i pæne, vestlige kredse, som startskuddet til, at det var skidt at ryge. Nøgternt set ved vi jo nok at der var nogen i Hitlers Tyskland, der fandt ud af det før, men det har man sådan ligesom glemt, for det var ikke helt comme-il-faut dengang. Så ja, det er jo så et eksempel på en undersøgelse, som starter vores viden omkring de negative effekter ved tobak, og derudover metodemæssigt så var det et godt studie med de forbehold han selv sidenhen fandt ud af. Så jeg mener klinisk men også rent metodemæssigt, så er det et landmark-studie.

Question: *De grundlægger vel nærmest et helt nyt forskningsfelt?*

Response: Ja, det er nok egentlig rigtigt. For historien, i hvert fald sådan som jeg har læst den går jo at den unge mand på det tidspunkt, Richard Doll, blev kaldt ind på chefens kontor og fik at vide, “Gå ud, og find ud af hvorfor vi har en 15-fold stigning af lungekræfttilfælde her i de sidste 10-15 år.” Og så vidt jeg er orienteret, så var han selv ryger på det tidspunkt. Det holdt han så op med. Og det er jo så også en BMJ-sag.

Question: *Når du så nævner case-control studier, er det så et spørgsmål om, at de samtidig laver metodeudvikling?*

Response: Jamen det har han jo gjort.

Question: *Altså han har udviklet en ny metode?*

Response: Mjah, nu kommer jeg jo en lille smule til kort, fordi han har formentlig været med til at udvikle metoden, men der er formentlig andre, epidemiologer, som har forestået selve metodeudviklingen. Jeg mener Jørn Olsen plejer at snakke om en anden, som en af teoretikerne bag ved det. Men Richard Doll han var simpelthen så snedig, at han godt kunne se, at der jo nok var et principielt problem i et case-control studie, så derfor dannede den her - resultaterne fra hans case-control studie - dannede baggrund for en af verdens længst fulgte kohorter. Det gav nemlig anledning til at han sendte brev ud til de her 35,000 engelske læger, der var engang i start 50'erne, og spurgte "ryger du/ryger du ikke?", og så i øvrigt fulgte dem de næste 50 år frem. Så i hvert fald forfatteren har været en foregangsmand i studiedesigns. Både det retrospektive case-control design og dernæst vende bøtten rundt og tage det i et kohortedesign. Han døde så for 6-7 år siden, efter at have fulgt dem i 50 år. Og det er det stort set kun Frammingham der har gjort ret meget længere. Det startede vel i '48, tror jeg nok. Og han har jo så i kohortedesign fulgt de her 35,000 engelske læger, og han har afrapporteret dem ca. hvert 10. år, og hans 50-års opfølgning den kom jo så her engang omkring årtusindskiftet, med et optryk af hans oprindelige artikel, som cadeau fra BMJ's side.

Question: *Ved (1)'eren snakkede du om, at det er sjældent man opfatter kasuistikker som høj kvalitets forskning. Hvordan med case-control studier? Hvordan er de i den forbindelse?*

Response: Jamen sådan på en almindelig rating, så har vi jo meta-analyser som top, så kommer randomiserede studier, så kommer kohorte, så kommer case-control lige efter. Og begrundelsen for at case-control studiet måske ligger lidt under kohorte, da er det væsentligste jo, hvordan man har defineret kontrolgrupperne. Har man skåret nogle hjørner der? Og det er der flere eksempler i historien på at man har fået fuldstændig misvisende fund ved at få valgt en forkert kontrolgruppe. Og der mener jeg faktisk at den gode Doll, plus andre af de helt top... Walter Willett, han har faktisk været med i en undersøgelse som de offentliggjorde i Lancet, hvor et blodtrykssænkende middel som hed.. navnet er væk nu, var forbundet med forøget risiko for kræft, så nu skulle alle folk gå hjem og holde op med at bruge det. Og så var der nogen der kom til at tænke bagefter at de havde fået valgt deres kontrolgruppe helt forkert. Og da man så fik en ny kontrolgruppe ind fandt man ud af, at det var helt forkert, den oprindelige konklusion. Så selv der i starten af 70'erne er der topfolk indenfor den her epidemiologiske branche som er gået i vandet. Det er jo sjovt nok.

Question: *Når jeg nu siger kvalitet, og du så begynder at snakke om evidens, betyder det så at der er en sammenhæng imellem forskningskvalitet og evidensniveau?*

Response: Jaeh, er der ikke det? Det tror jeg nok jeg vil sige. Men selvfølgelig, forskning kan være mange ting.

Question: *Og du siger selvfølgelig også selv, at det kan godt være god kvalitet selvom det er en kasuistik.*

Response: Altså, det her fokuserer jo ret meget på risiko, prognose og behandling. Og så kan der være mange andre forskningsområder, som jeg ikke kender så meget til. Men jo, jeg vil nok mene at evidens og kvalitet, det må hænge sammen.

Question: *Hvad med den tredje?*

Response: Det er jo en superartikel. Den er om 4S-studiet. Før den tid var det sådan, at der var en heftig debat. Den var faktisk perfid og ubehagelig, f.eks. i læserbreve i Ugeskrift for Læger, hvor de svinede hinanden til, de folk der var overlæger på kardiologisk afdeling i Aarhus. Nogle mente jo det bare var en association, det med at kolesterol og hjertekarsygdomme havde noget med hinanden at gøre, andre påstod det var en kausalitet. Og 4S-studiet er så det første studie der once-for-all i et randomiseret design viser at der er faktisk en kaustisk sammenhæng imellem kolesterol og dødelig hjertesygdom. Den er skandinavisk, og jeg mener der må have været flere andre studier undervejs, for der komme flere ganske kort tid efter, udenlandske studier, men de her var de første. Så jeg har altid brugt den her, som noget folk indenfor vores branche burde vide, ligesom Alexander Flemming opfandt Penicilin og sådan nogle ting. Og det er jo så et randomiseret studie, det er jo et andet studiedesign.

Question: *Hvis jeg nu bare siger ordet forskningskvalitet, hvad tænker du så på? Hvis man nu i stedet for de her tre artikler specifikt, skal prøve at snakke mere generelt om det?*

Response: Hvad jeg tænker? Jamen det er jo noget i retning af, at hvis der skulle være noget kvalitet i det, så skulle det være nogenlunde stringent tænkt, og det skulle være nogenlunde stringent udført. Og jo da fra min stol også gerne klinisk relevant, for en større gruppe personer. For der kan jo være elegante designs som vedrører en eller anden minoritet, hvor jeg måske nok vil sige det var fin kvalitet, men altså, det er så fjernt fra min virkelighed, og så lille en gruppe, at jeg ville tænke "Fair nok, at der er nogen der vil beskæftige sig med det lille hjørne, men det vil ikke få nogen impact for den samlede sundhed."

Question: *Så man kan godt skelne imellem kvaliteten af det enkelte studie, og så kvaliteten i forhold til impact, altså et gennemslag i den generelle sundhed?*

Response: Ja, i høj grad. Der kan jo være meget elegante og også vældig banebrydende studier, som omfatter en eller anden patientgruppe som har f.eks. 100 patienter i Danmark. Det er selvfølgelig vældig godt, jeg læste lige, for nu at tage et eksempel, at for folk med hudcancer blev der frigivet et stof, som koster en vældig masse penge, og som de regner med vil kunne udskyde en 6-7 dødsfald om året i Danmark. Der ligger der givetvis nogle vældig fine randomiserede studier til grund for det, ikke fordi jeg har læst dem, men det må jeg gå ud fra at der gør, siden man har valgt at anbefale den behan-

dling man gør. Men det vil ikke kunne rokke, antydningvis, ved den gennemsnitlige levealder her. Det vil være de fem personer, der vil have gevinsten ved det. Og helt i den omvendte gade, der er der nogle af de nye stoffer til blodfortyndingsbehandling ved atrieflimren, det vil jo få impact for 50-60,000, ikke nødvendigvis for deres livslængde men for deres livskvalitet, må man gå ud fra.

Question: *Så for mig lyder det lidt som om du siger, at der er en kvalitet som går på hvordan studiet er designet og udført, og så er der en kvalitet der går på hvor stor betydning det har.*

Response: Ja, men det er måske snarere min personlige kæphest. For det ene er validiteten af den dokumentation der er, og det andet er jo så den praktiske implikation af det. Og personligt synes jeg jo det er fint nok, der er nogen der arbejder med de der ganske sjældne ting.

Question: *Hvad så med grundforskning?*

Response: Jaeh, det, det må folk da gerne gøre for min skyld.

Question: *Men det er ikke noget du personligt opfatter som så afgørende?*

Response: Nåmmen det kommer da nok af at jeg er rundet af noget praktisk. Jeg ville aldrig nogensinde kunne holde til at bruge mit liv på noget, som nogen måske en gang om føje år vil kunne bruge til et eller andet, hvor der ikke er nogen nyttegevinst ved det lige nu. Og hvis du er tilpas perfid, så kan du jo spørge hvad nytte det så har at se på hvordan det går børn født af mødre med sukkersyge, men sådan har vi jo så hver vores små nicher, for det er da heller ikke fordi det berører så voldsomt mange. Men ren grundforskning, nej, det ville jeg ikke gide. Men derfor er der selvfølgelig nogen der skal gøre det, for det ender jo nok på et eller andet tidspunkt med, at der er nogen der kan få noget ud af det. En del af det i hvert fald.

Question: *Hvis man nu skal forsøge at evaluere forskning, f.eks. forsøge at afgøre hvilket universitet der har et større eller bedre output end et andet. Er der noget man kan gøre der? Er der noget du mener man kan gøre?*

Response: Det var et godt spørgsmål. Altså, jeg vil da indrømme, at når jeg ser de her voluminøse forskningsberetning, som jo blandt andet det her hus excellerer i, men det gør de da også i Aarhus og alle mulige andre steder, føler jeg godt nok nogle gange at det bliver blæst noget op, i den grad.

Question: *Hvordan tænker du?*

Response: Når du ser nogle afdelinger her, og du ved hvor lidt de laver, og så har de en eller anden mand, som engang har været ansat på klinisk Epidemiologisk Afdeling, eller hvad det nu kan være, eller måske noget der bliver lavet på et helt andet sygehus, så

ser man på deres hjemmeside, eller de rapporter I udgiver, at så er den afdeling sandelig også kommet med på det. Jeg erindrer da på et tidspunkt for nogle år siden, hvis vi nu skal sætte navn på, at Gynækologisk Afdeling havde tre af deres væsentligste arbejder, de var udgået fra Klinisk Epidemiologisk Afdeling, og så var der et eller andet i noget gynækologi og obstetrik obskurt blad, som jeg aldrig havde hørt om, men altså, hvad hedder hun nu, Mettes, det var i et pænt blad.

Question: *Men hun var vel ansat på afdelingen?*

Response: Hun fik ikke een krone i støtte fra Gynækologisk/Obstetrisk Afdeling, no nothing. Og hvis vi skal gå endnu mere personligt ind i det, så en af de artikler hun var med til, det var en hvor Margrethe Møller var med som medforfatter; hun havde aldrig rørt en finger, det var i Diabetes Care. Der spurgte jeg nemlig Hans ad, om ikke han kunne støtte lidt, men nej, hans penge var ikke til sådan noget. Men han kunne da være med til at betale paper charge, de der 3,000 kroner det kostede at få det publiceret i Diabetologia. Og ved senere at snakke med dem derinde, der fik jeg at vide, at støtten kunne da være, at de havde en professor der ville sætte navn under en ansøgning til nogle fonde. Men var det ikke egentlig langt væk fra spørgsmålet?

Question: *Joeh, jeg kan faktisk ikke lige huske hvad det var... jo... hvorvidt man kan evaluere forskning?*

Response: Jamen jeg bliver dig egentlig svar skyldig. Det med bare at tælle artikler, jeg synes det kan blive noget vældig misvisende noget. Det synes jeg. Og du har jo hele diskussionen her Aalborg og Aarhus imellem, om de her 500 artikler som Klinisk Epidemiologisk Afdeling har lagt navn til. Og der er... nej, det synes jeg er svært. Jeg ved ikke om de der citationsfaktorer kan bidrage med noget, men det ved du jo også selv, og har selv nævnt, at det kan være tvivlsomt.

Question: *Er der sammenhæng imellem forskningsartikler, og så den forskning der ligger bag?*

Response: Den skal jeg lige have en gang til.

Question: *Hvis du sidder med en artikel, som er rigtig god, bygger den så nødvendigvis på et godt studie, eller kan du godt sidde med en god artikel, der bygger på et dårligt studie?*

Response: Nej, det tror jeg ikke man kan. Hvis man nu er dygtig nok til engelsk, så kunne jeg måske nyde det...

Question: *Jo jo, men indholdsmæssigt?*

Response: Narh, det tror jeg, jeg vil sige nej til.

Question: *Så hvis man kunne evaluere kvaliteten af artiklerne, så ville man faktisk også ad omveje have en evaluering af selve forskningen?*

Response: Ja, det ville man nok. Altså, hovedpointen ved alt det der peer review, det er jo trods alt, med alle de fejl og mangler der er, at der er nogen der har gennemskuet de værste faldgruber og udeladelser. Men omvendt så vil jeg da indrømme, at jeg har ladet mig ophidse over studier, som jeg mener har været oversolgt i en helt ekstrem grad. Selv i BMJ, det skuffer mig dybt. Så det kan da godt være jeg bliver nødt til at modificere en lille smule... men nej, det må jo nok være sådan et nogenlunde fast ja til at der er en sammenhæng.

Question: *Kunne man opfatte det som et særligt kvalitetsstempel, at en undersøgelse blev brugt som en reference i en klinisk guideline?*

Response: Ja, det tror jeg egentlig godt. For der må man jo gå ud fra, at der er nogle folk der har siddet og læst det, og ment det var en kvalitet som kunne danne baggrund for en praksis i en afdeling, så det vil jeg gå ud fra. Og så kan du jo selvfølgelig sige, jamen de sidder i de der advisory boards og de får penge for det. Yes, det kan godt være, men alligevel, det må nok betragtes alt i alt som et ekstra kvalitetsstempel.

Question: *Ja... ved du tilfældigvis hvor ofte de her tre artikler er citeret?*

Response: Nej, det ved jeg ikke.

Question: *Det er bare fint, det forlanger jeg heller ikke at du ved. Det er egentlig mest for at vide om du havde kigget inden du valgte dem.*

Response: Nej, overhovedet ikke.

Question: *Det sidste jeg har er en brainstorm, hvor jeg nævner nogle begreber, og så vil jeg gerne høre din umiddelbare reaktion på det. Selvfølgelig i konteksten forskningskvalitet, og du bestemmer selv hvor meget du vil sige. Det første begreb jeg har, det er layout af artikler.*

Response: Det er folk blevet langt bedre til. og indenfor visse grænser er det egentlig godt, for det gør det mere appetitvækkende. BMJ gør det jo sådan ret populært på en eller anden facon, men så har de jo deres hele artikler, som man kan gå ind på nettet og finde. Det kan da også overdrives. Altså, jeg ved da at Ugeskriftet vil have et billede til hver artikel, det er måske lige på overdrevet. Men det er nok alt i alt positivt at der er kommet mere fokus på layout, det gør det mere appetitligt.

Question: *Hvad med sådan noget som sproget? Er det en afgørende faktor for at man vurderer en artikel som høj eller måske middelhøj kvalitet?*

Response: Den skal jeg lige have igen... tænker du på om det er engelsk eller dansk?

Question: *Nej, jeg tænker egentlig mere på de sproglige evner ved forfatteren?*

Response: Jamen der tror jeg egentlig jeg kan fortælle dig en historie, som du ikke har hørt før, måske. Jeg fik engang en artikel i BMJ, det var mig der var førsteforfatter, det var ikke så meget mig der havde skrevet den, jeg udtænkte planerne. Den kom i BMJ, og jeg havde jo så noget korrespondence med BMJ. Vi havde haft en, fordi vi erkendte vores begrænsede evner indenfor engelsk, vi havde haft den oppe ved John Andersson, som var sprogværter, og vi havde haft den frem og tilbage flere gange, og han er sådan british english, og ganske ferm til engelsk, og han havde været den igennem flere gange. Så blev den accepteret, og så kom den tilbage fra BMJ efter de havde haft den til sprogrevison. Der var ikke een eneste sætning der stod, som jeg havde skrevet den. Den var fuldstændig ændret...

Question: *... og det var efter I havde haft den til revision?*

Response: Ja. Og der var nogle få steder, hvor de havde misset pointen, det var de kvikke nok til at rette ret hurtigt. Så jeg må sige, alle BMJ's artikler er fuldstændig omskrevet til den style de gerne vil have, så forfatterens egen stil kan man ikke nødvendigvis genkende. Jeg var dybt forbavset. Men det er klart, det giver en ensartethed i det sproglige. Jeg går så ud fra, at der ikke er så mange, der gør det så heftigt, som de gjorde der. Men jeg mener egentlig at på den baggrund, så kan det være lidt svært, for det kan godt være, at det bare er nogle sprogdaktører du læser. Og ikke nødvendigvis forfatterne selv. Og en anden ting, så skrap er jeg ikke til engelsk at jeg kan afgøre om det er... ja, jeg kan se om der er nogle formuleringer, som jeg synes er smarte, men nej, det tør jeg ikke at sige. Så til dit spørgsmål, helt kontant, vil jeg sige nej.

Question: *Okay, godt. Tværfaglighed?*

Response: Mener du imellem forskellige personalegrupper?

Question: *Det kunne det være, ja, det kunne også være specialer, eller fag.*

Response: Tværfaglighed imellem fag er nok et absolut gode, for de vil hver især kunne bibringe noget. Om det er ... ja, jeg tror egentlig bare jeg vil stoppe der.

Question: *Okay. Ny viden?*

Response: Jamen det er jo nødvendigt. Er det ikke det vi er her for?

Question: *Det er der mange der mener, i hvert fald.*

Response: Ja ja, sådan i mine yngre dage, så kunne jeg godt finde på at spørge, når folk kom og sagde, at de ville et eller andet, "er det for dit CV eller for livet eller verden?", og jeg kan da huske at jeg en gang til noget ph.d.-undervisning i Aarhus, at jeg i kaffepausen

spurgte de her unge, håbefulde mennesker, hvor mange artikler de regnede med at få ud af det her, jaeh, det var jo nok en tre-fire stykker per person. Hvor mange af de her artikler de så regnede med førte til noget? Så blev de jo lidt mere lange i spyttet, så bare for det hold vil der jo være en små hundrede artikler, der kan komme ud og forurene PubMed, uden at de vil bringe nogensomhelst nytte. Så det var jo nok mest deres CV det skulle bringe videre. Men det er nok bare en gammel mands refleksioner.

Question: *Hvad med sådan noget som impact?*

Response: Jamen hvis det er klinisk impact i relation til at det bliver brugt til en klinisk vejledning, så er det jo fint. Men du har engang fortalt mig at LIFE var en af de mest citerede danske studier. 40,000 eller sådan noget, var det ikke det du sagde?

Question: *Det kan jeg ikke huske, men det skal nok passe. Det er i hvert fald meget højt citeret.*

Response: Det kunne jeg altså godt finde på at sige, at det er i hvert fald oversolgt i en helt ekstrem grad. SÅ hvis impact er det samme som citationer, så mener jeg altså ikke nødvendigvis, at sund sans og klinisk relevans hænger sammen med impact. Jeg ved ikke om... nej, det har du ikke været med til, men jeg har nogle gange i de der artikellæsningsseancer taget to eksempler, dels LIFE, det er noget med 9,000 mennesker der er trukket lod til den ene slags blodtryksnedsættende og den anden slags og så følger man dem over fire år, og så ser man så, at der er en lillebitte forskel, og qua man har puttet en hulens masse folk ind kan man gøre det statistisk signifikant. Og der er noget med at man skal behandle adskillige mange hundrede for at opnå eet effektmål. Det er jo sådan det ene studie, og det andet eksempel, det er så Ole Bonderup og kumpaner heroppe fra, der tog tyve med kollagent colit, og gav dem bhudenosid og et eller andet placebo i den anden og så fulgte dem i seks uger, og fandt ud af, at de allesammen fik det bedre, hvis de fik det der bhudenosid. Hvad er stærkest af de to studier? Det her husmandsforsøg, tyve mennesker, fulgt over seks uger? Eller 10,000 mennesker, der er blevet screenet, 8,000 der er involveret og fulgt over det meste af verden over fem år? Oles er ikke så citeret, det er helt sikkert, men LIFE er citeret, det er helt sikkert.

Question: *Ja, og hvad havde egentlig størst effekt?*

Response: Jeg vil jo mene, at studiets styrke afhænger af antallet af udfald, og de fik allesammen et positivt udfald i Oles og de positive udfald hvad enten man fik beta-blokkere eller, nej, et af de to ting... der var langt mellem udfaldene.

Question: *Ja... hvad med sådan noget som peer review?*

Response: Ja, jeg er jo skidesur lige nu, for jeg har jo fået sådan en åndssvag artikel retur. Det er faktisk ret pudsigt, for der var tre reviewers, og den ene, det var kun kommaføj, det var en super artikel, den skulle absolut og aldeles publiceres. Og de to andre de var

“dunk dunk”, og det var sikkert nok nogle reviewere... de har jo formentlig taget dem, jeg selv havde valgt. Jeg skulle give nogle navne på nogle reviewere. Så de to: tomten nedad og den tredje tomten opad, og så total afvisning. Så dem kan jeg da godt have det lidt skidt med indimellem. Men... omvendt må vi jo erkende, at der var også nogle budskaber der var fornuftige blandt de der kritikere. Så jeg skulle egentlig have siddet, hvis ikke der var et andet møde i eftermiddag, og rettet noget af det der til. Såeh...

Question: *Er det vigtigt vi har det?*

Response: Ja, det tror jeg det er. Jeg hørte Reimar fortælle, der var nogle af de anæstesiologiske blade, som egentlig havde en hel masse reviewers som bare fik en uge til at kigge det igennem. Så de havde egentlig ikke regnet med, at de allesammen skulle bruge en halv dag på at grave i dybden. Det var i hvert fald det han sagde.

Question: *Det tror jeg ikke helt jeg forstår.*

Response: Joeh, han sagde noget om at det blev sendt til flere, men de havde ikke nogen forventning om, at folk gravede så langt ned i substansen. Hvis du tager en artikel, jeg kan huske nogle af dem jeg har reviewet, der har jeg siddet og prøvet... brugt en halv dag på at komme igennem det, og dermed også haft lidt mere tid til det måske. Hvor Reimar sagde at nogle af de her anæstesiologiske tidsskrifter, de havde mange, langt flere end de her 1-2 reviewers.

Question: *Okay, så i stedet for en 2-3 grundige, så har de måske en 6-7 mere overfladiske...*

Response: Ja, det var i hvert fald det han sagde, jeg har ikke noget evidens for at det passer.

Question: *Ja, det er jo interessant i hvert fald, hvis der er nogen, der gør det på den her anden måde. Men det er sådan set det jeg har. Du må selvfølgelig gerne sige, hvis der er noget du sidder og tænker på, noget du synes du mangler at sige...*

Response: Det du vel er lidt ude efter er vel om jeg mener man kan foretage sådan en eller anden standardiseret vurdering af, om en afdeling eller institution eller sygehus eller hvad det nu er, er god eller dårlig til forskning...

Question: *Nej, det er det nu egentlig ikke. Men selvfølgelig, hvis du har nogle tanker omkring det, så vil jeg da gerne høre dem. Men det der er vigtigt for mig, det er egentlig at finde ud af, hvad I læger opfatter som kvalitet.*

Response: Jamen altså, behandlingsmæssigt, et randomiseret forsøg, det er stadig det eneste der kan sikkert fordele alle de der confounders, kendte og ukendte, målbare og ikke-målbare ved lodtrækning.

Question: *Men er de andre forsøg så overflødige i forhold til behandling?*

Response: Jamen altså, nu er du jo så sprogligt forkert på den, for er det forsøg, så bør det være lodtrækningsforsøg, altså ren intervention. Så er der de observerende designs, dem har jeg det vældig svært ved. Jeg plejer at bruge en artikel, som jeg også kunne have taget med, nemlig HERS, den undersøgelse, hvor man tog 1,400 kvinder, som havde haft blodprop i hjertet og randomiserede dem til østrogen og ikke-østrogen, og fandt ud af at det ikke hjalp spor, eller at man døde lidt mere, hvis man fik østrogen. Og det kom jo i halen af en forfærdelig masse observerende undersøgelser, der allesammen entydigt pegede på, at det var godt at spise det her østrogen. Og problemet, er man jo så sidenhen blevet klar over, at man jo optimistisk havde overfortolket de observerende undersøgelser, at man ikke lige havde lagt mærke til, at de stort set allesammen byggede på det der amerikanske Nurses Health Study, og for at du bliver sygeplejerske, og for at du gider som sygeplejerske at besvare det her skema, så er du allerede røget langt ned i tobaksforbrug, og dit BMI er faldet ganske betydeligt og dine motionsvaner er steget. Så det var jo en ikke-repræsentativ gruppe, som man tolkede meget bastant. Der var en lederartikel i det blad der, som skrev, at nu skulle man ikke længere tænke på, om kvinder efter overgangsalderen skulle have hormon, nej, nu skal vi tænke over, om der er undergrupper hvor vi skal sige, nej, de skal lade være med at få det. Eller kun have det i en periode. Og jeg mener egentlig det samme problem kan være for noget af det Mette Nørgaard går rundt og gør nu her, at kvinder, der har spist de her lykkepiller i løbet af en graviditet de også har nogle flere misdannelser. Konklusionen på alt det her er, de observerende undersøgelser, hvis man vil tolke at en eller anden sammenhæng man ser i en observerende undersøgelse den er kausal, så er man på tynd is i det midnste. Fordi der ikke er nogen der siger, at det kun er lykkepiller der er årsagen. Der kan være meget andet i det sociale; hvem er det, der får lykkepiller? Det er nogen der nok ryger lidt mere, det er nogen der nok har en lidt dårligere økonomi, de bor i Aalborg Øst eller hvad de nu gør, de har haft en hel masse forskellige andre sociale ting. Derimod synes jeg det er super når sådan en som Christial Melgaard kigger på "er der en sammenhæng imellem, om folk får autisme, når de har fået MFR-vaccine?". Og der tog de jo et skideelegant design, hvor de tog de ca. 500,000 børn, der er blevet vaccineret med den her MFR-vaccine, og så de 80-90,000 hvor forældrene har valgt ikke at få dem vaccineret, og så kiggede de på hvor mange der ender med at få en autisme-diagnose, og der var ikke en skid forskel.

Question: *Og det har man jo ellers haft en teori om, at der var.*

Response: Ja, lige præcis, så altså, det observerende studie, der ender med en nul-sammenhæng, det synes jeg er glimrende. Men det observerende studie, hvor man postulerer at et eller andet, man tilfældigvis har fundet som eksponering, det også giver et outcome, det er jeg noget mere loren ved.

Question: *Fordi det kun er associationer man kan finde?*

Response: Præcis. Og man skal være enormt forsigtig når man tolker fra associationer i observerende design til at det også er kausalt.

Question: *Men det kan vel være med til at føre hen til at finde noget kausalt?*

Response: Ja ja, lige præcis, ja, det kan det. Sådan lige for at springe i det, mit postulat kunne jo så være, at de kvinder, der har fået noget mere medicin, de også bliver observeret mere omhyggeligt, deres børn bliver observeret mere omhyggeligt, for at se om der skulle være et eller andet. Jeg har kigget på de børn som har haft hjertemisdannelse, kigget på grupperne hjertemisdannelse plus mor sukkersyge, hjertemisdannelse plus mor ikke sukkersyge, der synes jeg det var bemærkelsesværdigt, at dem der var i den gruppe der var hjertemisdannelse plus mor sukkersyge, der levede alle børnene, men i den anden gruppe var der otte der var døde. Det kunne jo tyde på, at hvis man som rask mor får et barn med hjertemisdannelse, så er de noget mere syge end hvis du som diabetesmor får et barn, hvor de lytter og kigger, fordi de ved de burde finde en eller anden misdannelse. Så finder de nogle mindre, mere subtile ting, som egentlige ikke quo ad vitam betyder så meget. Altså, på en eller anden led kan man jo sige, at det at få en diagnose i landspatientregistret, det er bare et proxymål for at man er syg. Det er ikke nødvendigvis det samme som at man er alvorligt syg, og du behøver ikke være lige alvorligt syg, med samme diagnose.

Question: *Så det er et spørgsmål om at man simpelthen bare overvåger mere?*

Response: Ja, det er surveillance-bias. Jeg har ikke fået det offentliggjort endnu, der er så mange tekniske... men det er også kun lige her for nylig at jeg fik tallene lavet. Men det... hvis du nu skal sammenligne med andre der sidder her, på et senere tidspunkt, så tror jeg da nok at en af følgerne af at jeg har været lidt blandet ind i det her epidemiologi det er at min skeptiske sans er blevet styrket.

References, main study

- Baschat, A. A. (2011). Neurodevelopment following fetal growth restriction and its relationship with antepartum parameters of placental dysfunction. *Ultrasound in Obstetrics and Gynecology*, *37*, 501-514.
- Black, D. M., Cummings, S. R., Karpf, D. B., Cauley, J. A., Thompson, D. E., Nevitt, M. C., ... Ensrud, K. E. (1996). Randomised trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. Fracture Intervention Trial Research Group. *Lancet*, *348*(9041), 1535-1541.
- Doll, R., & Hill, A. B. (1950). Smoking and carcinoma of the lung; preliminary report. *British Medical Journal*, *2*(4682), 739-748.
- Eissa, D., Carton, E. G., & Buggy, D. J. (2010). Anaesthetic management of patients with severe sepsis. *British Journal of Anaesthesia*, *105*(6), 734-743.
- Febbraio, M., Guy, E., & Silverstein, R. L. (2004). Stem cell transplantation reveal that absence of macrophage CD36 is protective against atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *24*, 2333-2338.

- Fischer, R. A., Nucci, M. R., Thaker, H. M., Weremowicz, S., Genest, D. R., & Castrillon, D. H. (2004). Complete hydatidiform mole retaining a chromosome 11 of maternal origin: molecular genetic analysis of a case. *Modern Pathology*, *17*, 1155-1160.
- Gedikli, O., Kiris, A., Ozturk, S., Baltaci, D., Karaman, K., Durmus, I., ... Celik, S. (2010). Effects of prehypertension on arterial stiffness and wave reflections. *Clinical and Experimental Hypertension*, *32*, 84-89.
- Greenhalgh, T. (2010). *How to read a paper: the basics of evidence-based medicine* (4th ed.). John Wiley and Sons.
- Hoberman, A., Paradise, J. L., Rocklette, H. E., Shaikh, N., Wald, E. R., Kearney, D. H., ... Barbadora, K. A. (2011). Treatment of acute otitis media in children under 2 years of age. *The New England Journal of Medicine*, *364*(2), 105-115.
- Hvid-Jensen, F., Pedersen, L., Drewes, A. M., Sørensen, H. T., & Funch-Jensen, P. (2011). Incidence of adenocarcinoma among patients with barrett's esophagus. *New England Journal of Medicine*, *365*(15), 1375-1383.
- Klatskin, G., & Kimberg, D. V. (1969). Recurrent hepatitis attributable to halothane sensitization in an anesthetist. *The New England Journal of Medicine*, *280*(10), 515-522.
- Koutcherov, Y., Mai, J. K., Ashwell, K. W., & Paxinos, G. (2000). Organization of the human paraventricular hypothalamic nucleus. *Journal of Comparative Neurology*, *423*(2), 299-318.
- Laurberg, P., Pedersen, K. M., Vestergaard, H., & Sigurdsson, G. (1991). High incidence of multinodular toxic goitre in the elderly population in a low iodine intake area vs. high incidence of Graves' disease in the young in a high iodine intake area: comparative surveys of thyrotoxicosis epidemiology in East-Jutland Denmark and Iceland. *Journal of Internal Medicine*, *229*(5), 415-420.
- Laurent, S., & Boutouyrie, P. (2007). Recent advances in arterial stiffness and wave reflection in human hypertension. *Hypertension*, *49*, 1202-1206.
- Liaw, Y. F., Sung, J. J. Y., Chow, W. C., Farrell, G., Lee, C. Z., Yuen, H., ... Cirrhosis Asian Lamivudine Multicentre Study Group (2004). Lamivudine for patients with chronic hepatitis B and advanced liver disease. *The New England Journal of Medicine*, *351*(15), 1521-1531.
- Livet, J., Weissman, T. A., Kang, H., Draft, R. W., Lu, J., Bennis, R. A., ... Lichtman, J. W. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, *450*(7166), 56-62.
- Lohse, N., Hansen, A. E., Pedersen, G., Kronborg, G., Gerstoft, J., Sørensen, H. T., ... Obel, N. (2007). Survival of persons with and without HIV infection in denmark, 1995-2005. *Annals of Internal Medicine*, *146*, 87-95.
- Mayberg, H. S., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., ... Kennedy, S. H. (2005). Deep brain stimulation for treatment-resistant depression. *Neuron*, *45*(5), 651-660.
- Mueller, S. (2008). Classification of eosinophilic gastrointestinal diseases. *Best Practice & Research Clinical Gastroenterology*, *22*(3), 425-440.
- Neziri, A. Y., Curatolo, M., Nüesch, E., Scaramozzino, P., Andersen, O. K., Arendt-Nielsen, L., & Jüni, P. (2011). Factor analysis of responses to thermal, electrical, and mechanical painful stimuli supports the importance of multi-modal pain assessment. *Pain*, *152*, 1146-1155.
- Olesen, S. S., Bouwense, S. A. W., Wilder-Smith, O. H. G., Van Goor, H., & Drewes, A. M. (2011). Pregabalin reduces pain in patients with chronic pancreatitis in a randomized, controlled trial. *Gastroenterology*, *141*, 536-543.
- Podrez, E. A., Byzova, T. V., Febbraio, M., Salomon, R. G., Ma, Y., Valiyaveetil, M., ... Hazen, S. L. (2007). Platelet CD36 links hyperlipidemia, oxidant stress and a prothrombotic phenotype. *Nature Medicine*, *13*(9), 1086-1095.
- Rivers, E., Nguyen, B., Havstad, S., Ressler, J., Muzzin, A., Knoblich, B., ... Early Goal-Directed Therapy Collaborative Group (2001). Early goal-directed therapy in the treatment of severe sepsis and septic shock. *New England Journal of Medicine*, *345*(19), 1368-1377.
- Rüschoff, J., Dietel, M., Baretton, G., Arbogast, S., Walch, A., Monges, G., ... Höpfner, H. (2010). HER2 diagnostics in gastric cancer - guideline validation and development of standardized im-

- munohistochemical testing. *Virchows Archiv*, 457, 299-307.
- Scandinavian Simvastatin Survival Study Group. (1994). Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the scandinavian simvastatin survival study (4s). *Lancet*, 344(8934), 1383-1389.
- Slim, R., & Mehio, A. (2007). The genetics of hydatidiform moles: new lights on an ancient disease. *Clinical Genetics*, 71(1), 25-34.
- The GRIT study group. (2004). Infant wellbeing at 2 years of age in the growth restriction intervention trial (grit): Multicentred randomised controlled trial. *The Lancet*, 364, 513-520.
- Thomsen, R. W., Schoonen, M., Farkas, D. K., Riis, A., Jacobsen, J., Fryzek, J. P., & Sørensen, H. T. (2009). Risk for hospital contact with infection in patients with splenectomy. *Annals of Internal Medicine*, 151, 546-555.
- Tähtinen, P. A., Laine, M. K., Huovinen, P., Jalava, J., Ruuskanen, O., & Ruohola, A. (2011). A placebo-controlled trial of antimicrobial treatment for acute otitis media. *The New England Journal of Medicine*, 364(2), 116-126.
- von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P., & the STROBE Initiative. (2007). The strengthening the reporting of observational studies in epidemiology (strobe) statement: Guidelines for reporting observational studies. *Annals of Internal Medicine*, 147(8), 573-577.
- Wiegand, J., Wedemeyer, H., Finger, A., Heidrich, B., Rosenau, J., Michel, G., ... Tillmann, H. L. (2008). A decline in hepatitis B virus surface antigen (HBsAg) predicts clearance, but does not correlate with quantitative HBeAg or HBV DNA levels. *Antiviral Therapy*, 13, 547-554.
- Zitvogel, L., Kepp, O., & Kroemer, G. (2011). Immune parameters affecting the efficacy of chemotherapeutic regimens. *Nature Reviews CLinical Oncology*, 8, 151-160.

B. Interview invitation

B.1. Original letter of invitation

The original letter can be seen on the following pages. A translated version is presented subsequently.

«Adressekartotek»

PILOTPROJEKT
(også ikke-læger inviteres til at deltage)

«Hilsen»

Mange tak for at Du vil deltage i denne undersøgelse, til anvendelse i mit ph.d.-projekt. Undersøgelsen fokuserer på lægers opfattelse af forskningskvalitet, især i form af videnskabelige artikler og lignende publikationer. Undersøgelsen har form af et 30-minutters interview, hvor jeg vil stille dig nogle spørgsmål omkring forskningskvalitet. Der vil ikke være andre til stede under interviewet.

For at give Dig bedst mulighed for at vælge et sted og tidspunkt, hvor du har mulighed for at deltage, bedes Du bruge oversigten på denne hjemmeside:

<http://ahsic.dk/rq/book>

Du skal bruge følgende informationer for at logge på siden:

Brugernavn: «ID»

Adgangskode: «Kode»

Til undersøgelsen bedes Du venligst medbringe følgende:

- Underskrevet samtykkeerklæring (vedlagt dette brev).
- 2-3 videnskabelige artikler, som Du vurderer som værende af særligt høj kvalitet. Der er intet krav til artiklernes emne, aktualitet eller form. Dog vil jeg bede dig om at vælge artikler på engelsk eller dansk.
- Udfyldt formular med baggrundsinformationer (vedlagt dette brev, men kan også udfyldes på ovennævnte hjemmeside).

Med venlig hilsen

Jens Peter Andersen

Medicinsk Bibliotek

Aalborg Sygehus

Forskningens Hus
Sdr. Skovvej 15
9000 Aalborg
Tlf.: 9932 6823

www.aalborgsygehus.rn.dk

Ph.d. studerende
Jens Peter Andersen
Direkte: 9932 6990
jepea@rn.dk

Ref: «ID»

Samtykkeerklæring

Til anvendelse ved undersøgelsen: "Lægers opfattelse af forskningskvalitet"

Deltager: «Fornavn» «Efternavn»

Ref.: «ID»

Undertegnede giver hermed samtykke til:

- At informationer, meninger og andre udsagn afgivet i forbindelse med interview til ovennævnte undersøgelse kan anvendes og formidles af ph.d. studerende Jens Peter Andersen, i forbindelse med dennes ph.d. studium.
- At informationer, der ikke er af personfølsom art må genbruges i fremtidige forskningsprojekter.

Undersøgelsens ophav, ph.d. studerende Jens Peter Andersen, garanterer til gengæld for dit samtykke:

- At alle personlige oplysninger krypteres eller maskeres, så der ikke kan drages sammenhæng imellem din person og afgivne informationer.
- At ingen afgivne informationer på nogen måde videresælges.
- At ingen personlige informationer på nogen måde videregives.

Underskrift

Baggrundsinformationer

Ref.: «ID»

Stilling _____

Forskningsaktivitet Ingen
 Tidligere
 Aktiv

Forskningserfaring Ingen
 Nogen
 Meget

Uddannelse Medicinstuderende
 Læge
 Ph.d.-studerende
 Ph.d.
 Speciallæge
 Dr. Med.
 Andet: _____

Speciale _____

Alder _____

Køn Mand
 Kvinde

B.2. Translated letter of invitation

NB! Only the main letter text is presented here. The declaration of consent is available in appendix C.3.

Thank You for participating in this study, which is a part of my PhD project. The study focuses on medical doctors' perception of research quality, especially in the shape of scientific articles and similar publications. The investigation is a 30 minute interview, during which I will pose questions about research quality to you. Nobody else will be present during the interview.

To offer You the best possible choice of location and time of your participation, please use the overview on this website:

<http://ahsic.dk/rq/book>

You need the following information to log in to the site:

Username: «ID »

Password: «code »

For the interview I would kindly ask You to bring:

- Signed declaration of consent (attached to this letter).
- Two or three research articles, which You consider to be of particularly high quality. There are no requirements of topic, news value or form. However, I would ask you to select articles in English or Danish.
- Filled form with background information (attached to this letter, but can also be filled out on the above website).

Best wishes,
Jens Peter Andersen

C. Interview documents

C.1. Interview script, pilot study

Q: Thank you for participating in this study. During the interview, I will ask you to answer some questions regarding your perception of research quality. The interview will be recorded on a dictaphone, but your responses will be anonymised in all further use.

* Dictaphone is turned on.

Q: You have brought some articles, which you consider especially high quality. For reference, I will number them.

* Articles are numbered.

Q: Would you say something about why you chose these particular articles?

Q: Is there something concrete about the individual articles which you would emphasise as being decisive for their quality?

Q: I would like to know your thoughts about the phrase research quality.

Q: If one were to assess the quality of research, where would you consider the focus should be?

Q: With regard to quality, how do research publications relate to the research behind them?

Q: I will mention some phrases, and I would like you to briefly respond with the first

thoughts that come to mind:

- New knowledge
- Cross-disciplinarity
- Impact
- Citations
- Layout of articles
- Peer review

Q: Thank you for your participation.

C.2. Interview script, main study

Q: Thank you for participating in this study. During the interview, I will ask you to answer some questions regarding your perception of research quality. The interview will be recorded on a dictaphone, but your responses will be anonymised in all further use.

* Dictaphone is turned on.

Q: This is interview with reference «ID ». You have brought some articles, which you consider especially high quality. For reference, I will number them.

* Articles are numbered.

Q: Would you say something about why you chose these particular articles?

Q: Is there something concrete about the individual articles which you would emphasise as being decisive for their quality?

Q: I would like to know your thoughts about the phrase research quality.

Q: If one were to assess the quality of research, where would you consider the focus should be?

Q: With regard to quality, how do research publications relate to the research behind them?

Q: What kind of connection is there between level of evidence and research quality?

Q: Is there a feeding chain for articles, e.g. from case studies, via randomised trials to guidelines?

Q: Are you aware of how many times the articles you brought have been cited?

Q: I will mention some phrases, and I would like you to briefly respond with the first thoughts that come to mind:

- New knowledge
- Cross-disciplinarity
- Impact
- Citations
- Layout of articles
- Peer review

Q: Thank you for your participation.

C.3. Declaration of consent

NB! This declaration was originally in Danish and is presented here in English.

Declaration of Consent: For the investigation “Medical doctors perception of research quality”

Participant: «first name »«last name »

Ref.: «ID »

I, the undersigned, hereby declare my consent to:

- informations, opinions and other statements declared in connection with interviews for the abovementioned study can be used and mediated by PhD student Jens Peter Andersen, in connection with their PhD study.
- non-personal information may be reused in future research projects.

The investigator, PhD student Jens Peter Andersen, guarantees in return of the consent, that:

- all personal information is encrypted or masked, so that no connection can be drawn between the person and given information.

Conceptualising research quality in medicine for evaluative bibliometrics

- no information will ever be sold to third party.
- no personal information will ever be given to third party.

Signed:

D. Initial statement codes

Table D.1: Initial statement codes and unique ID's as well as frequency (f) of occurrence

ID	Statement code	<i>f</i>
Ethics in research		
141	Ethical research	5
Factors related to the author(s)		
1	Authors experience as researcher	2
2	Authors famousity in the field	4
54	Some research is conducted mainly to appear on CV's	2
Factors related to the impact of articles		
106	High-impact journals have more valid research (more strictly peer-reviewed)	3
88	Journal indicates article quality	1
59	Discrepancy between journal impact and article quality	4
39	Low-impact journals publish high-quality articles, but may be more likely to also include low-quality articles	3
33	High-impact journals are built on news-articles, not necessarily high-quality research	1
161	Fraud is more likely in high impact journals	1
118	High impact factor is a self-feeding mechanism - high impact factors attract better papers	1
146	Papers in high-impact journals are more likely to be cited	3
168	High impact journals may be less likely to publish groundbreaking research, as it may be outside main-stream	1

Continued on next page

Table D.1 – continued from previous page

ID	Statement code	<i>f</i>
80	There is a bigger incentive to commit fraud, if trying to publish in the Lancet	1
34	When submitting, high impact factor is important, as it is used for grant applications	5
170	Clinical=high impact, basic=low impact	1
137	Impact is field specific	1
123	High-impact journals tend to prefer (high-evidence) clinical articles	3
68	High impact factor is indicative of the quality of articles	9
102	Information overflow leads to higher discriminability of high- and low-impact journals	1
142	Impact factor is overrated	1
37	Impact indicates utility, not quality	1
Clinical practice guidelines as specific document type		
134	References in clinical guidelines are usually of especially high quality	1
149	Clinical guidelines are arcane (old boys around the table, what we think we know)	1
Other specific study types		
81	Systematic reviews are essential for clinical evidence	1
150	Discoveries begin with case studies	1
105	Proper systematic reviews are necessary, not only to show the evidence, but also to point out what research needs to be done	1
138	All researchers should do systematic reviews	1
82	Systematic reviews indicate high quality	1
Groundbreakingness of research		
93	Genious is art	1
92	Research is a skill (trainable), genious is chance and circumstance	1
129	Groundbreaking/novel ideas are more important than novel methods	1
115	High evidence does not mean groundbreaking	3
100	Groundbreaking research is of high quality	1
25	Groundbreaking research gives more citations	4
29	Groundbreaking research may be slow to publish	1
Citations		
73	Citations are given to popular articles (fashion-effect)	1
119	Citation scores are used when seeking information - highly cited papers may be more interesting	1

Continued on next page

Table D.1 – continued from previous page

ID	Statement code	<i>f</i>
132	Citations as indicator of quality are doubtful	1
124	Citations reflect mainstream articles	1
120	Citations are important for grant applications	1
165	Selfcitations can be self-promoting	1
164	Selfcitations may be necessary	1
24	If a paper is cited, it is more likely to be cited again (because of laziness - reference list copying)	4
58	Citations are contextual (discrepancy between specialties)	3
147	"old" researchers/authors tend to be more cited, as "old" authors cite each other (Cronyism/old boys around the table)	1
70	Citations are more relevant for evaluation than journal impact factors	7
32	Citation impact indicates degree of groundbreaking, quality is a matter of the research behind	4
74	High citation counts indicate quality	7
145	Big names are more cited	1
152	There is a connection between citations and quality, but it is not straightforward	3
96	Citations indicate centrality	2
26	Citations are given to high-profiled names	1
References		
78	References are given as reward	1
48	Relates to previous research	1
50	Correct use of references (references should be relevant)	5
166	More likely to use previously cited work (signals merit)	1
Evidence		
114	There are other qualities than evidence	1
52	Evidence is not just randomised trials, it may be multimodality (basic research evidence)	1
83	As a clinician, you need high evidence papers (trials or above)	2
67	Evidence levels depend on the type of research/hypothesis	1
Research as a source of inspiration		
3	Research as source of inspiration	1
Interdisciplinarity		
7	Basic research requires interdisciplinarity with clinicians	2

Continued on next page

Table D.1 – continued from previous page

ID	Statement code	<i>f</i>
126	Researchers should seek information outside of their field	1
28	Interdisciplinarity may give novel ideas (synergy-effect)	15
55	Interdisciplinarity is not a goal in itself	1
71	Interdisciplinarity is important	3
Methodology and materials		
157	Connection between problem type and study type	3
128	Robustness	1
77	Study design	1
103	Correct selection criteria (data)	1
56	Papers may be a result of repetitive work on datasets (e.g. for student projects)	1
97	Quality depends on planning, structuring and executing research stringently	1
45	Certain methods may be golden standards	2
111	Check for bias	1
158	Validity	3
11	Reproducibility	27
44	Multimodality	7
10	Strong methodology	17
76	Study design	3
101	Research needs to be free of methodological constraints - freedom to use qualitative methods in a quantitative world	1
Factors related to journals		
127	Editorials might point to particularly good papers	1
89	Overflow of journals	1
Novelty		
14	The paper founded a new approach	1
19	New knowledge is a product of research	14
61	New knowledge is the aim of research	9
5	Novelty has relevance	10
15	The paper succesfully introduces a new treatment	1
16	Novelty is a necessity (information overflow)	1
53	Research that doesn't generate new knowledge is irrelevant	1
18	Introduces a new method	3
36	If research produces new knowledge, it is a quality	1
135	Originality [novelty]	1

Continued on next page

Table D.1 – continued from previous page

ID	Statement code	<i>f</i>
Results		
30	Method matches hypothesis	3
35	Quality of materials	1
31	Reporting of results must be precise and unbiased	1
169	Longitudinal data	1
133	All results must be reproduced	3
64	Validity of documentation	5
49	Statistical significance	1
6	Negative findings may be interesting and important	4
21	Sizeable material	18
162	Well executed studies (e.g. stopping when critical mass is reached)	1
153	Prospectivity is a quality	1
41	Economic use of data; to extract all possible knowledge from data	1
42	Economic use of data; to conclude as soon as possible (not collect too much data)	1
167	Data	1
62	Reproducing results is good, but at some point it becomes wasteful	2
90	Unimodal analyses of large datasets may be useful	1
23	Study design	2
Style (e.g. visual/textual)		
131	Good papers should have structured, informative abstracts	1
143	Stringency	1
116	(good) language is important	9
13	Visual layout of papers is important for the perception of a paper and therefore the quality	13
98	Thoroughness	1
79	Scientific structure (layout) is essential	1
99	Method section (layout) is irrelevant, noone reads it	1
148	Planning	3
66	Structure	1
86	Adherence to publishing guidelines	1
104	Online-only journals may be of higher quality, as they don't limit articles (room for proper methods etc.)	1
Basic and clinical research		
22	Researchers should also be clinicians	2

Continued on next page

Table D.1 – continued from previous page

ID	Statement code	<i>f</i>
113	Basic research is distanced from the clinical life	1
125	Basic research may be of little relevance in clinical work	1
63	Free, independent, basic research is important for new development and groundbreaking research	1
94	Clinical work is part relying on research and part craftsmanship	1
122	Basic research is usually not published in high-impact journals	1
151	Clinical changes come from basic research	3
85	Results should be possible to realise/implement	3
87	Clinically relevant results should be short and to the point	1
108	Research should be substantial	1
47	The ultimate purpose of basic (medical) research is to help people	1
65	Practical/clinical implications	3
51	Basic research should not focus on patients, but on how the world is connected	1
136	Clinical relevance is highly important	2
154	The path from basic to clinical research is so long, it is not apparent	1
144	Clinical research should strive for great health impact	1
139	RCTs are good for the clinic. For research, other study types may work better.	1
20	Clinicians need to do research to be able to relate to the research literature	4
163	The path from basic to clinical research is not straight, rather zig-zag	3
140	Basic research should lead to clinical research	5
72	Basic research creates the difference	8
40	In a clinical setting, clinical relevance is a quality indicator	8
4	Clinical research builds on basic research	10
Implementation / translation		
38	Bad research may generate new hypotheses, good research may become implemented	1
Peer review		
95	Peer reviewing may be flawed - it depends on the skill/-knowledge of the referees	1

Continued on next page

Table D.1 – continued from previous page

ID	Statement code	<i>f</i>
69	Peer review is of higher quality in high impact journals	3
60	Peer review is a gatekeeper	7
8	Peer review is important	20
117	Statistical knowledge is required to assess single papers	1
Research environment		
130	A good research group promotes an environment that lets e.g. grant students continue to research after conclusion of PhD's	1
Clinical relevance		
9	Usefulness in clinical setting	5
27	Research papers can be used as a supplement to hand-books	1
112	Clinical relevance and research relevance are very different	1
110	Clinical impact depends on the number of outcomes	1
121	Relevant hypothesis	1
109	Research should be worthwhile	1
156	Local research fitting local conditions	1
159	Clinical impact may be geographically dependant, which may affect citation impact (high clinical impact in africa is not likely to get high citation impact as opposed to US)	1
43	Research should present information that is useful for people	1
160	How many people does it affect?	1
155	Research should give actual results, and not just point out the need for more research	1
46	Research should serve a direct purpose	3
Various connections between other factors		
107	Connection between evidence and quality	16
84	Research should lead to improved health	1
57	Clinical impact and citation impact should be related	1
91	Connection between citations and evidence level	1
12	Quality and evidence level are related	8
17	Connection between quality of papers and quality of research	29

E. Initial survey questions

E.1. Personal details

Please provide the following information. All details will be kept confident and anonymous.

1. Year of birth
2. Country of residence
3. Years of research experience

E.2. Statements on research quality

The following statements have been collected from interviews with medical doctors and all relate in some way to research quality. Please indicate on a scale from 1-5 to what degree you agree, with 1 = strongly agree, 2 = agree somewhat, 3 = indifferent, 4 = somewhat disagree, 5 = strongly disagree.

Author factors

1. Papers written by experienced authors are generally of higher quality
2. Authors' renown may affect the perceived quality of their papers
3. Some research is mainly conducted to appear on CVs or to fulfill research quotas.

Journal Impact Factor

1. High-impact journals are more strictly peer-reviewed than other journals

2. Journal impact factor gives an indication of the quality of papers published in the journal.
3. Low-impact journals are more likely to include low-quality papers than high-impact journals.
4. Low-impact journals also publish high-quality papers.
5. High journal impact factors are an incentive for selecting a journal for publishing a paper.
6. Scientific fraud is more likely in high-impact journals.
7. High-impact journals focus on mainstream papers.
8. When selecting references for my own papers, I am more likely to select ones from high-impact journals.
9. Journal impact factor is important when applying for grants.
10. Clinical research is more likely to be published in high-impact journals than biomedical or basic research is.
11. Journal impact factor indicates the usefulness of research articles.

Clinical guidelines

1. References used in clinical guidelines are usually of especially high quality.
2. Clinical guidelines are essential for implementing research in practice.
3. Clinical guidelines are based on what we think we know, rather than evidence.

Ground-breaking research

1. Ground-breaking research is in a league of its own with qualities beyond normal research.
2. Research can be considered a skill or craftsmanship, while genius is art, chance or circumstance.
3. Ground-breaking research is not necessarily on a high evidence level.
4. Ground-breaking research is generally slower to publish than mainstream research.
5. Ground-breaking research gives more citations.

Citations

1. There is a connection between citations and the quality of a research paper.
2. Mainstream papers receive more citations.
3. I am more likely to use the work of my colleagues for references, as I know their work better.
4. I am more likely to use the work of a well-known researcher in my field, as their work is generally respected.
5. I am more likely to use highly cited articles for references.
6. Citations tell you more about the quality of the paper than the journal impact factor of the journal in which it was published.

Evidence

1. Multimodal methods give higher evidence.
2. I am more likely to give references to studies on a high evidence level.
3. For clinicians, randomised controlled studies, meta-analyses and systematic reviews are most useful.
4. For researchers, studies need not have high evidence to be useful.
5. Systematic reviews are essential to provide evidence for quality in practice.
6. Discoveries begin with case studies.

Basic and clinical research

1. Basic research is required for ground-breaking research to exist.
2. Changes in clinical practice originate from basic research.
3. There is a gap between basic research and clinical work.
4. Research should be substantial.
5. The ultimate purpose of basic research is improved health.
6. Research should be clinically relevant.
7. Basic research focuses on how the world is connected, not on patients.

Clinical practice

1. Research papers can be used as a supplement to handbooks.
2. Articles may be relevant for other research or the clinic, but not necessarily both.
3. Research should give actual results, and not just point out the need for more research.

Peer review

1. Peer reviewing may be flawed, as it depends on the skill/knowledge of the referees.
2. Peer review excludes the lowest quality research.
3. Peer review guarantees high quality research.

Journals

1. If an article is mentioned in the editorial, it is usually a particularly good article.
2. There is an overflow of journals.
3. Online-only journals may be of a higher quality as print-journals, as they don't limit articles, giving more room to methods, discussion etc.

Results

1. Negative findings may be interesting and important.
2. Reproducing results is good, but at some point it becomes wasteful.
3. Unimodal analyses of large datasets may be useful.

Interdisciplinarity

1. Basic research requires interdisciplinarity with clinicians.
2. Researchers should seek information outside their field.
3. Interdisciplinarity may give novel ideas (synergy-effect).
4. Interdisciplinarity is a goal in itself.

Methodology/data

1. Quality depends on planning, structuring and executing research stringently.
2. Certain methods are golden standards, and should be used when possible.
3. Research needs to be free of methodological constraints, allowing medical researchers to use qualitative methods or alternative methods.

General relationships

1. There is a positive connection between evidence level and research quality.
2. Research should lead to improved health.
3. There is a positive connection between clinical impact and citation counts.
4. There is a positive connection between citation counts and evidence level.
5. There is a positive connection between the quality of research and the quality of articles.

E.3. Importance of article-specific factors

Please rate the following items on a scale of 1-5, indicating their importance with respect to the quality of a research article. 1 = highly important and 5 = unimportant.

Novelty

1. The paper presents a new method.
2. The paper presents new knowledge on a subject.
3. The paper introduces a new treatment.

Materials

1. Longitudinal data.
2. The material involves a large number of cases (patients, events, samples etc.)
3. Prospective data.
4. Economic use of data; don't collect too much data and extract all potential knowledge from collected data.
5. Correct selection criteria.
6. Quality of materials.

Methods

1. The applied method matches the hypothesis of the paper.
2. The study type matches the problem.
3. Robustness.
4. Multimodality.
5. Strong methodology.

Results

1. Validity.
2. Check for bias.
3. Precision.
4. Reproducibility.
5. Statistical significance.
6. The results may be directly implementable.
7. Clinically relevant results.
8. Number of outcomes.
9. Local research fitting local conditions.

Style

1. Structured, informative abstract.
2. Stringency; keeping it to the point.
3. Language-skill.
4. Visual layout.
5. Thoroughness in description of methods, materials and results.
6. Scientific structure in the layout.
7. Method section.
8. Review section.

References

1. Relates to previous research.
2. Correct use of references; found the most relevant references for the article.

Other

1. Peer review.
2. Generates new hypotheses.
3. Usefulness in a clinical setting.
4. Potential effect on patients (number of people, size of impact).
5. Adherence to ethical guidelines.
6. Interdisciplinary research.

F. Final survey

The visual design of the survey is shown first, followed by detailed response options for drop-down fields.

F.1. Visual design

F.2. Drop-down response options

Primary job function

- Research
- Clinical practice/Physician/Surgeon
- Administration
- Management
- Other

Specialty

- Accident and emergency medicine
- Allergy and Immunology
- Anaesthetics
- Anatomy
- Audiological Medicine



Research Quality Survey "Medical researchers' perception of research quality"

Take the survey

To get started on the survey, please enter your email in the field below, and click "submit". For confirmation purposes, please use your **institutional email** (university, hospital, company etc.).

Confidentiality & cookies

Your email is used for verification purposes, and to allow you to leave the survey and come back later. We will never relay your contact information to other parties, or give access hereto.

When you enter the survey, a cookie^[2] is stored on your device with your login-ID. No further information is stored, and the cookie is automatically deleted after 60 minutes.

Need more information?

You can read a brief outline of the study in this PDF: [Study outline \(PDF, 174 KiB\)](#)

Contact information

The survey is conducted by PhD student Jens Peter Andersen as a collaboration between the Medical Library at Aalborg Hospital, Aarhus University Hospital and the Royal School of Library and Information Science.

Contact the surveyor at:

E-mail: rq@ahsic.dk

Tel: (+45) 9932 6990

Postal: Medical Library, Sdr. Skovvej 15, DK-9000 Aalborg, Denmark

AALBORG HOSPITAL
AARHUS UNIVERSITY HOSPITAL



IVA
Royal School
of Library and
Information
Science



Research Quality Survey
"Medical researchers' perception of research quality"

1. Personal details 2. Statements on research quality 3. Importance of article-specific factors 4. Additional comments

Part 1. Personal details

Please provide the following information. All details will be kept confident and anonymous.


Year of birth	Select... ▼
Gender	Select... ▼
Years since completed formal medical education	0
Primary job function	Select... ▼
Specialty	Select... ▼
Primary work place	Select... ▼
Approximate number of publications where you are (one of) the author(s)	0
Country of residence	Select... ▼
Please check this, if you would like to receive a report on the results	<input type="checkbox"/>

[submit answers](#)

AALBORG HOSPITAL
AARHUS UNIVERSITY HOSPITAL



IVA Royal School
of Library and
Information
Science



Research Quality Survey
"Medical researchers' perception of research quality"

1. Personal details
2. Statements on research quality
3. Importance of article-specific factors
4. Additional comments

Part 2. Statements on research quality

In the following questions, please indicate to what degree you agree with the proposed statements.

Please indicate to what degree you believe the author(s) of an article to be decisive for how you experience the quality of an article

0 = not at all, 10 = very decisive

0 5 10

○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○

Articles written by experienced authors are generally of higher quality

1 = disagree completely, 5 = agree completely

1 5

○ ○ ○ ○ ○

If I know an author as a good researcher, I am inclined to perceive their new articles to be of high quality before reading them

1 5

○ ○ ○ ○ ○

"I find there are more research journals than necessary", do you agree with this statement?

0 = disagree completely, 10 = agree completely

0 5 10

○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○

Many journals limit the length of articles. Do you believe this to be:

1 = disagree completely, 5 = agree completely

- A problem, if the methods and discussion sections cannot be elaborated

1 5

○ ○ ○ ○ ○

- A good thing, as there is little time for reading lengthy articles, and they must therefore be short and concise

1 5

○ ○ ○ ○ ○

If an article is mentioned in the editorial of a journal, it stands out from the other articles in that issue.	<p>1 = disagree completely, 5 = agree completely</p> <p>1 . . . 5</p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>
Some journals generally publish articles of higher quality than other journals.	<p>0 = disagree completely, 10 = agree completely</p> <p>0 5 10</p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>
Please indicate on a scale whether your primary source for <i>keeping up to date</i> on new research is journals (e.g. subscriptions, e-alerts, table of contents, RSS) or by searching databases (e.g. PubMed, Embase)	<p>Please indicate on a scale where 0 = journals exclusively, 10 = databases exclusively</p> <p>0 5 10</p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>
The journal impact factor is indicative of the quality of articles appearing in a journal	<p>0 = disagree completely, 10 = agree completely</p> <p>0 5 10</p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>
High impact factors are an incentive when I select a journal to publish a paper in	<p>1 = disagree completely, 5 = agree completely</p> <p>1 . . . 5</p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>
To what degree do you agree with this statement: "When selecting references for my own papers, I am more likely to select ones from high-impact journals"	<p>0 = disagree completely, 10 = agree completely</p> <p>0 5 10</p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>
The quality of an article is independent of the journal it appears in	<p>0 5 10</p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>
To what degree do you agree with this statement: "Clinical guidelines reflect the highest quality research available on a topic"	<p>0 = disagree completely, 10 = agree completely</p> <p>0 5 10</p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>
References used in clinical guidelines are usually of especially high quality	<p>1 = disagree completely, 5 = agree completely</p> <p>1 . . . 5</p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>
Clinical guidelines are decisive for implementing research in practice	<p>1 . . . 5</p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>
Clinical guidelines are based on what we think we know, rather than evidence	<p>1 . . . 5</p> <p><input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/></p>

Conceptualising research quality in medicine for evaluative bibliometrics

If an article is cited by many other articles, I would consider this to indicate: *0 = disagree completely, 10 = agree completely*

- High quality research	0 5 10 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
- Recognition / collegial respect	0 5 10 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
- The article is on a hot topic	0 5 10 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

I am more likely to use highly cited articles as references	<i>1 = disagree completely, 5 = agree completely</i> 1 5 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
I am more likely to cite the work of a well-known researcher in my field	1 5 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
I am more likely to cite the work of my colleagues	1 5 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

Some research can be considered ground-breaking, which is an outstanding quality	<i>0 = disagree completely, 10 = agree completely</i> 0 5 10 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Changes in clinical practice can be led back to advances in basic research	0 5 10 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
The ultimate purpose of basic research is improved health	0 5 10 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

Research - whether basic or clinical - should be clinically relevant	<i>1 = disagree completely, 5 = agree completely</i> 1 5 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Peer review is a gatekeeper, assuring a minimum of quality in published research	1 5 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
More strict peer review would lead to higher quality in published research	1 5 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Peer review may prevent new and innovative or groundbreaking ideas from being published	1 5 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Negative findings may be interesting and important	1 5 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Certain methods are golden standards and should be used whenever possible	1 5 <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

Part 3. Importance of article-specific factors

Below are five topics relating to the quality of journal articles. For each topic please select the three elements you would consider most decisive for the quality of an article in descending order.

Materials

Selection 1	<input type="text" value="Select..."/>
Selection 2	<input type="text" value="Select..."/>
Selection 3	<input type="text" value="Select..."/>

Methods

Selection 1	<input type="text" value="Select..."/>
Selection 2	<input type="text" value="Select..."/>
Selection 3	<input type="text" value="Select..."/>

Results

Selection 1	<input type="text" value="Select..."/>
Selection 2	<input type="text" value="Select..."/>
Selection 3	<input type="text" value="Select..."/>

Style

Selection 1	<input type="text" value="Select..."/>
Selection 2	<input type="text" value="Select..."/>
Selection 3	<input type="text" value="Select..."/>

Other

Selection 1	<input type="text" value="Select..."/>
Selection 2	<input type="text" value="Select..."/>
Selection 3	<input type="text" value="Select..."/>



Research Quality Survey
"Medical researchers' perception of research quality"

1. Personal details 2. Statements on research quality 3. Importance of article-specific factors **4. Additional comments**

Part 4. Additional comments

If you have any additional comments on the topic "research quality" in the field below. Feel free to include any topics you believe are missing, or if you want to elaborate on any of the topics so far. If you have no further comments, feel free to leave this field empty.

Comments:

We would like your permission to contact you in case we have follow-up questions. By checking this mark, you allow us to contact you and save your e-mail address for that purpose. In case we don't contact you, your e-mail will be removed from our system along with all other e-mails, once the survey has been finally concluded.

submit answers

AALBORG HOSPITAL
AARHUS UNIVERSITY HOSPITAL



IVA Royal School
of Library and
Information
Science

- Biochemistry
- Cardiology
- Cardiothoracic surgery
- Child and adolescent psychiatry and psychotherapy
- Clinical Genetics
- Clinical neurophysiology
- Clinical Pharmacology and Therapeutics
- Dental surgery
- Dermato-Venerology
- Dietetics
- Embryology
- Endocrinology
- Family Medicine
- Forensic medicine
- Gastroenterology
- General medicine
- General surgery
- Genetics
- Geriatrics
- Gynaecology and obstetrics
- Haematology
- Health informatics
- Infectious diseases
- Internal medicine
- Interventional radiology
- Laboratory medicine

- Maxillo-Facial surgery
- Microbiology
- Neonatology
- Nephrology
- Neurology
- Neuroradiology
- Neurosurgery
- Nuclear medicine
- Occupational medicine
- Oncology
- Ophthalmology
- Orthopaedics
- Otorhinolaryngology
- Paediatric surgery
- Paediatrics
- Palliative Medicine
- Pathology
- Pharmacology
- Physical and rehabilitation medicine
- Plastic, reconstructive and aesthetic surgery
- Pneumonology
- Podiatry
- Psychiatry
- Public health
- Radiology
- Radiotherapy

- Rehabilitation
- Rheumatology
- Sports and exercise medicine
- Thoracic surgery
- Toxicology
- Tropical medicine
- Urology
- Vascular medicine
- Vascular surgery

Primary work place

- General practice
- Private company (e.g. Medicinal)
- University
- University hospital/Teaching hospital
- Private hospital
- Public hospital
- Other public healthcare
- Other private healthcare

Materials

- The paper includes longitudinal data
- The material involves a large number of cases (patients, events, samples etc.)
- Data are prospective
- Economic use of data; extract all potential knowledge from collected data and don't collect too much data
- Correct selection criteria
- Quality of materials

Methods

- The applied method matches the hypothesis of the paper
- The study matches the problem
- Multimodality (use of several methods)
- Reproducibility
- Detailed descriptions of the methods
- Innovative methodology
- Strong/robust methodology

Results

- Validity
- Bias has been considered and avoided where possible
- Precision
- Statistical significance
- The results may be directly implementable in practice
- Clinically relevant results
- Number of outcomes
- Local research fitting local conditions

Style

- Structured, informative abstract
- Stringency; keeping it to the point
- Language-skill
- Visual layout
- Thoroughness in description of methods, materials and results
- Scientific structure in the layout (order of sections)
- Presence and content of the methods section
- Presence and content of the review section

Other

- Relates to previous research
- Correct use of references; found the most relevant references for the article
- Peer review
- Generates new hypotheses
- Usefulness in a clinical setting
- Potential effect on patients (number of people, size of impact)
- Adherence to ethical guidelines
- Interdisciplinary research

G. E-mail invitation to survey

Subject line: What is research quality? Your opinion counts Please take a moment to consider this invitation to my survey, as I believe it may be of interest to you and your research.

For your convenience, here is a very short explanation, the longer version can be seen below:

What: Survey on medical researchers perception of research quality

Why: To investigate the concept research quality for use in research assessment

Where: <http://rq.ahsic.dk>

How long: 15-20 minutes

Who: Jens Peter Andersen, PhD student

What I would like you to do: Answer a number of multiple-choice questions with your honest, personal opinion. You are not required to write long essays, but there will be an optional comments-field.

The long version: My name is Jens Peter Andersen, I am a PhD student at Aalborg Hospital and the Royal School of Library and Information Science in Denmark. I am conducting a survey on how research quality is perceived among medical researchers. The survey is based on interviews with a number of medical doctors, and what I need from you is your honest opinion on whether you agree with a range of statements.

You have been selected to participate because you recently were the corresponding author on a number of papers published in medical journals. In particular I noticed your paper titled XX which has been cited Y times - this is a count which is commonly employed in the assessment of research. Whether you agree on the use of citations to evaluate research impact or not, I would greatly appreciate your opinion on the mat-

ter. If you are interested, please go to <http://rq.ahsic.dk> to take the survey. The survey should not take more than 15-20 minutes.

The survey is a central element in my PhD dissertation in which I aim to conceptualise research quality, describe the dimensions of the concept and compare these to bibliometric methods and provide recommendations for a best-practice in research quality assessment.

A word on ethics and confidentiality: Your privacy is of great importance to me. Your email-address will never be passed on to a third party, nor will any connection between you as a person and the opinions you may express in the survey ever be published or passed on to a third party. When the survey is concluded, any personal information will be deleted from the dataset, and only non-personal identifiers will be used to handle data. Data will be discernable on a national level, but no closer than that, also keeping your institutions confidentiality intact.

Best regards,

Jens Peter Andersen

PhD student

Aalborg Hospital and Royal School of Library and Information Science

Aalborg, Denmark

email: rq@ahsic.dk

H. Data cleaning

Below is the source code for the R file cleaning the original survey data for input errors.

```
d <- read.table("input/survey.Rdat", header=TRUE, sep=";")

#REVISE EDUCATION VALUES
d$education.rev <- d$education
d$education.rev[d$education.rev == 1982] <- 2012 - 1982
d$education.rev[d$education.rev == 1996] <- 2012 - 1996
d$education.rev[d$education.rev == 1974] <- 2012 - 1974
d$education.rev[d$education.rev == 1992] <- 2012 - 1992
d$education.rev[d$education.rev == 1997] <- 2012 - 1997
d$education.rev[d$education.rev == 1977] <- 2012 - 1977
d$education.rev[d$education.rev == 1995] <- 2012 - 1995
d$education.rev[d$education.rev == 1998] <- 2012 - 1998
d$education.rev[d$education.rev == 2002] <- 2012 - 2002
d$education.rev[d$education.rev == "012"] <- 12
d$education.rev[d$education.rev == "036"] <- 36
d$education.rev[d$education.rev == "11_(technical!)"] <- 11
d$education.rev[d$education.rev == "-"] <- 0
d$education.rev[d$education.rev == "11_yrs"] <- 11
d$education.rev[d$education.rev == "24_(Midwifery_Education)"]
  <- 24
d$education.rev[d$education.rev == "7.5"] <- 7
d$education.rev[d$education.rev == "n.a."] <- 0
d$education.rev[d$education.rev == "n/a"] <- 0
d$education.rev[d$education.rev == "None"] <- 0
d$education.rev[d$education.rev == "not_medical,_PhD,_2003"]
  <- 0
d$education.rev[d$education.rev == "10_med_school/3_fellowship"]
  <- 10

#REVISE PUBLICATION COUNT VALUES
```

Conceptualising research quality in medicine for evaluative bibliometrics

```
d$pubcount.rev <- as.factor(d$pubcount)
# d$pubcount.rev[d$pubcount == "188 (excluding
  abstracts and letters)"] <- 188
# one field was manually corrected in the original data,
# as the above line did not work.
d$pubcount.rev[d$pubcount == ">150"] <- 150

# Assign correct data types
d$pubcount.rev <- as.numeric(as.character(d$pubcount.rev))
d$education.rev <- droplevels(d$education.rev)
d$education.rev <- as.numeric(as.character(d$education.rev))
```