Composite Measures of Quality of Healthcare

Kara, Pinar

DOI (link to publication from Publisher):
10.54337/aau510601530

Publication date:
2022

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

Link to publication from Aalborg University

Citation for published version (APA):
Kara, P. (2022). *Composite Measures of Quality of Healthcare*. Aalborg Universitetsforlag.
https://doi.org/10.54337/aau510601530

COMPOSITE MEASURES OF QUALITY OF HEALTHCARE

BY
PINAR KARA

DISSERTATION SUBMITTED 2022

AALBORG UNIVERSITY
DENMARK

COMPOSITE MEASURES OF QUALITY OF HEALTHCARE

by

PINAR KARA

AALBORG UNIVERSITY
DENMARK

Dissertation submitted 2022

.

Dissertation submitted: September 2022

PhD supervisor: Prof. Søren Paaske Johnsen, MD, PhD
Danish Center for Clinical Health Services Research,
Aalborg University

Assistant PhD supervisors: Prof. Jan Mainz, MD, PhD, MPA
Aalborg University

Jan Brink Valentin, MSc
Danish Center for Clinical Health Services Research,
Aalborg University

PhD committee: Clinical Professor Tarec Christoffer El-Galaly (chair)
Aalborg University, Denmark

Professor Marie Eriksson
Umeå University, Sweden

Professor Mette Nørgaard
Aarhus University, Denmark

PhD Series: Faculty of Medicine, Aalborg University

Department: Department of Clinical Medicine

ISSN (online): 2246-1302
ISBN (online): 978-87-7573-826-7

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Stibo Complete, 2022

ENGLISH SUMMARY

Measurement and quantification of the quality of healthcare is an essential step towards quality improvement. Individual quality indicators are important tools that provide valuable and detailed insights into the quality. However, the complexity and multi-dimensionality of the quality of care may require assessment of a large number of indicators to cover different aspects and dimensions of quality. Evaluation of the quality based on large number of indicators can be challenging and time consuming and may create a critical information burden for the users.

Composite measures, combination of multiple individual indicators, enable users to quantify overall quality of care with a single score. Use of composite indicators have gained attention in the recent years due to its simplicity when presenting and communicating quality of care with healthcare stakeholders, and its ability to demonstrate the changes in the quality of care over time and the variations in the quality of care between healthcare providers or systems with a single score.

This PhD thesis included three studies that covered different methodological aspects regarding construction, use and evaluation of composite measures. The thesis includes a review of the literature and two nationwide cohort studies based on Danish registries where different types of composite measures were constructed and ranking of hospitals based on different composite measures was compared. In addition, composite measures' association with the clinical outcome and the predictive accuracy in relation to the clinical outcome were examined.

In study I, a literature search was performed to investigate the use of composite measures to quantify the quality of healthcare services in the literature with an attention to methodological assessments in the reviewed studies. The identified studies were examined regarding (1) the used construction methods and (2) whether the constructed composite measures were subject to methodological considerations and whether these considerations (for example, potential limitations of composite scores) were communicated to the readers. Overall results from study I showed that the attention to methodological aspects of using composite measures was sparse.

In study II, the overall quality of care provided to patients with incident heart failure was quantified using different types of composite measures. Hospitals were ranked according to their composite scores, and the rankability of composite measures was investigated. Furthermore, the association between the composite measures and 1-year all-cause mortality was assessed.

In study III, overall process quality of care for patients with incident ischemic stroke was quantified using different types of composite measures, hospitals were ranked according to their performance based on composite scores, and rankability of

composite measures was investigated similar to study II. The association between the composite measures and 30-day mortality was assessed. Additionally, performance of composite measures in predicting 30-day all-cause mortality for incident ischemic stroke patients was examined.

Results from study II and study III demonstrated differences in rankings of hospitals based on different construction methods that cannot be fully attributed to random variation. This showed the importance of caution and awareness whenever using composite measures to compare healthcare providers as use of different methods led to change in performance rankings for some of the hospitals. In both of the studies, the majority of the composite measures were associated with reduced mortality. In study III, it was seen that all types of composite measures performed poorly regarding prediction of 30-day mortality for the patients with acute ischemic stroke.

The results from this thesis underlined that while use of composite measures can be seen as a simple solution to summarize quality of care, it requires (1) careful, step-by-step evaluation of the construction process and (2) clear and transparent communication of the construction process of composite measures and potential limitations associated with composite measures.

DANSK RESUME

Måling og kvantificering af kvaliteten i sundhedsvæsenet er en afgørende forudsætning for kvalitetsudvikling. Kvalitetsindikatorer er i den forbindelse værdifulde værktøjer, som kan give vigtig indsigt i kvaliteten. Sundhedsvæsenets kvalitet er dog kompleks og multidimensionel, hvilket ofte vil medføre at det er nødvendigt med et betydeligt antal forskellige individuelle kvalitetsindikatorer for at kunne belyse kvaliteten tilstrækkeligt. Det store antal kvalitetsindikatorer repræsenterer imidlertid også en udfordring i sig selv, idet det kan være vanskeligt og tidskrævende for modtageren at danne sig et overblik.

Kompositte mål, d.v.s. kombination af flere individuelle kvalitetsindikatorer, muliggør at der kan etableres en samlet værdi eller score for sundhedsvæsenets behandlingskvalitet på et givet område. Brug af kompositte mål har tiltrukket sig øget opmærksomhed i de senere år p.g.a. muligheden for på en enkel måde at præsentere og formidle kvaliteten i sundhedsvæsenet ved hjælp af én scoreværdi.

Denne PhD afhandling omfatter tre studier som omhandler forskellige metodologiske aspekter vedrørende konstruktion, brug og evaluering af kompositte mål, inklusiv en gennemgang af den eksisterende videnskabelige litteratur og to landsdækkende kohorte studier baseret på danske kliniske kvalitetsdatabaser, hvor forskellige typer af kompositte mål blev konstrueret og evalueret i forhold rangordning af hospitaler, associationen med kliniske udfald af patientforløb og prædiktive evne i forhold til kliniske udfald.

Studie I omfattede en systematisk litteratursøgning med henblik på at undersøge brugen af kompositte mål af kvaliteten i sundhedsvæsenet i den eksisterende videnskabelige litteratur med et særligt fokus på metodologiske forhold. De identificerede studier blev gennemgået i forhold til (1) hvilke typer af kompositte mål der blev anvendt og (2) hvorvidt de kompositte mål blev underkastet en kritisk vurdering samt, hvorvidt disse overvejelser blev formidlet til læserne, f.eks. om potentielle metodologiske svagheder. Generelt viste studie 1 at der i den eksisterende videnskabelige litteratur vedrørende kompositte mål for kvaliteten i sundhedsvæsenet kun har været begrænset opmærksomhed på metodologiske forhold.

I studie II blev behandlingskvaliteten blandt patienter med incident hjertesvigt vurderet ved hjælp af forskellige typer af kompositte mål. Hospitalerne blev rangeret i forhold til deres kompositte score og rangeringen ved anvendelse af forskellige typer af kompositte mål blev sammenlignet. Endvidere blev associationen mellem de forskellige typer af kompositte mål og 1 års mortalitet undersøgt.

I studie III blev behandlingskvaliteten blandt patienter med incident iskæmisk stroke vurderet ved hjælp af forskellige typer af kompositte mål svarende til

fremgangsmåden i studie II. Tilsvarende blev hospitalerne rangeret i forhold til deres komposite score og rangeringen ved anvendelse af forskellige typer af kompositte mål blev sammenlignet. Endelig blev associationen mellem de forskellige typer af kompositte mål og 30 dages mortalitet undersøgt ligesom de kompositte måls evne til at prædiktere 30 dages mortalitet blev beregnet.

Resultaterne fra studie II og III viste at rangeringen af hospitalerne varierede i forhold til hvilken type af komposit mål der blev anvendt og at denne variation ikke kunne forklares alene ved tilfældig variation. Dette understreger behovet for opmærksomhed og omhu når der anvendes kompositte mål til at sammenligne hospitaler eller andre kliniske enheder, idet forskellige typer af kompositte mål kan fører til forskellige konklusioner. I begge studier var de fleste typer af kompositte mål associeret med lavere mortalitet. Endvidere viste studie III at de kompositte mål kun i meget begrænset omfang kunne prædiktere 30 dages mortalitet hos patienter med iskæmisk stroke.

Resultaterne fra denne afhandling understreger at skønt kompositte mål kan være en simpel måde at sammenfatte omfattende datamængder vedrørende kvaliteten i sundhedsvæsenet, så forudsætter en rational anvendelse at der (1) foretages en omhyggelig trin-for trin gennemgang af, hvordan de kompositte mål er konstruerede og (2) en tydelig og transparent formidling af konstruktionsprocessen samt de metodologiske begrænsninger som knytter sig til de forskellige typer af kompositte mål.

LIST OF PAPERS

This PhD thesis was based on the following three papers:

PAPER I

Kara P, Valentin JB, Mainz J, Johnsen SP (2022) Composite measures of quality of health care: Evidence mapping of methodology and reporting. PLoS ONE 17(5): e0268320. doi: 10.1371/journal.pone.0268320. Published.

PAPER II

Kara P, Valentin JB, Løgstrup BB, Schjødt I, Egstrup K, Vinter N, Mainz J, Johnsen SP. Composite measures of quality of care for patients with heart failure in Denmark: A nationwide cohort study. In preparation.

PAPER 3

Kara P, Valentin JB, Damgaard D, Vögele P, Rasmussen TS, Murante AM, Vainieri M, Mainz J, Johnsen SP. Composite measures of quality of care for patients with incident ischemic stroke in Denmark: A nationwide cohort study. In preparation.

LIST OF ABBREVATIONS

AON	All-or-none
AUC	Area under curve
CI	Confidence interval
CRS	Civil registration system
DHFR	Danish Heart Failure Registry
DNPR	Danish National Patient Registry
DPR	Danish National Prescription Registry
DSR	Danish Stroke Registry
EW	Expert weights
FA	Factor analysis
HF	Heart failure
HR	Hazard ratio
LVEF	Left ventricular ejection fraction
NYHA	New York Heart Association
OBCS	Opportunity-based composite score
OECD	Organisation for Economic Co-operation and Development
OR	Odds ratio
PCA	Principal component analysis
ROC	Receiver operating characteristic
RKKP	The Danish Clinical Quality Program, National Clinical Registries
WHO	World Health Organisation

LIST OF TABLES AND FIGURES

Tables

Figures

ACKNOWLEDGEMENTS

This thesis originates from three studies conducted during my PhD at Danish Center for Clinical Health Services Research, Aalborg University and Region North Jutland and was fully funded by European Commission.

My deepest grattidute goes to my main supervisor, Søren Paaske Johnsen, for his great mentorship, trust, support, patience, kindness and for finding time whenever needed although his busy schedule. Pursuing a PhD can be challenging, even more so in a period with a pandemic and multiple lockdowns, but I've always felt Søren's support and he provided a solution for every challenge on the way. My gratitude also goes to my co-supervisors; Jan Mainz for his support, continuous feedback and encouragement and to Jan Brink Valentin for his inspiringly-fast solutions for every statistical challenge, and for sharing his expertise and providing feedback.

A special thanks to the funder of the studies in this thesis, the European Commission via Marie Skłodowska–Curie H2020 research and innovation funding programme ("HealthPros"- Healthcare Performance Intelligence Professionals, under grant agreement no. 765141) which made this PhD possible.

I feel very blessed to be a part of the HealthPros team. I'd like to thank all HealthPros fellows, PIs, support staff and consortium. A special thanks to HealthPros fellows for their interdisciplinary feedback, support and friendship. I feel incredibly lucky to be able to work with such bright people, for our network meetings every six months, being able to travel and have social activities together. They made my PhD years exceptional.

Thanks to the Department of Psychiatry, Region North Jutland and the Department of Clinical Medicine and International Staff Unit at Aalborg University. A special thanks to Mette Munk for helping me with everything, starting from my very first day in Denmark.

Thanks to all colleagues at Danish Center for Clinical Health Services Research (DACS) for creating a stimulating and pleasant work environment, for all intellectual discussions, and also for all the cakes and social activities.

Thanks to all co-authors. To Inge Schjødt, Kenneth Egstrup and Brian Bridal Løgstrup for heart failure care expert panel and to Dorte Damgaard, Peter Vögele and Tine Steenholt Rasmussen for stroke care expert panel. Their contrubutions brought clinician perspective into study II and study III. Thanks to Nicklas Vinter for sharing his extensive knowledge on heart failure and heart failure registry and providing detailed feedback. Thanks to Anna Maria Murante and Milena Vainieri for collaborating on study III, providing feedback and also making me feel very welcomed in Sant'Anna School of Advanced Studies in Pisa, Italy during my short research visit.

Thanks to Nihal Ata Tutkun, my master's degree supervisor who inspired me deeply for research and to pursue a PhD.

And last but the most important, I would like to express my greatest appreciation for my family. I am eternally thankful for my parents, Sevgi and Hasan Kara, and their unconditional love, trust, support and sacrifices.

Pinar Kara

Aalborg, 2022

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

Measurement and quantification of quality is a key step towards improvement of quality in healthcare. [1] Quality measurement is needed in order to track progress of healthcare systems, to identify weaknesses in a healthcare system, to assess whether current healthcare policies are effective and working optimally, and to take action according to the obtained information. [2]

While individual quality measures are important tools for quality improvement, increasing number of quality measures and increasing volume of information may create a challenging burden for users of this information. [3] Consequently, more information may not always lead to more actionable results.

Composite measures or indicators are tools which summarizes information from multiple measures and dimensions into one single score. [4, 5]These measures have the advantage of facilitating a simpler presentation of the quality, which may ease communication with the public and facilitate tracking of progress of healthcare providers and healthcare systems, as well as assessment of the effectiveness of quality improvement programmes, guidelines, policies, etc. [3-5]

While these measures have received increasing attention in the recent years due to their advantages over individual indicators, they are also associated with important potential limitations. [3-6] Hence, it has been emphasized that composite measures' usefulness, validity and quality may highly depend on the underlying data as well as the adopted approach for the construction process of the specific composite measures. [5] Therefore, it is timely and needed to examine the effects of using different composite scoring methodologies on results (for example, hospital rankings) and to address potential challenges with these measures.

CHAPTER 2. AIMS

2.1. OVERALL AIMS

The overall aim of this thesis was to demonstrate the use and interpretation of composite measures and to investigate the strengths and limitations associated with these measures when they are used to summarize the quality of care.

2.2. PAPER I

The main aims of study I were as follows,

I. To identify the methods that were used in the literature to construct composite measures of quality of care based on process indicators.

II. To investigate methodological considerations behind these methods: (1) if authors provided any justification regarding their choice of methodology to create composite measures, (2) if authors provided any information on advantages or limitations of using composite measures and (3) if authors informed the reader about the presence of other approaches to construct composite measures.

2.3. PAPER II

The main aims of study II were as follows,

I. To rank hospitals using multiple types of composite measures of incident heart failure care based on various weighting and aggregation schemes and to investigate how hospitals' rankings change according to the method that was used to construct composite measures.

II. To investigate the association between different types of composite measures and 1-year all-cause mortality for patients with incident heart failure in Denmark.

2.4. PAPER III

The main aims of study III were as follows,

I. To rank hospitals using different types of composite measures of incident ischemic stroke care based on various weighting and aggregation schemes and to investigate how hospitals' rankings change according to the method that was used to construct composite measures.

II. To investigate the association between different types of composite measures and 30-day all-cause mortality for patients with incident ischemic stroke.

III. To investigate the predictive accuracy of different types of composite measures in relation to 30-day all-cause mortality for patients with incident ischemic stroke in Denmark.

CHAPTER 3. BACKGROUND

3.1. QUALITY OF CARE

Quality of care can be defined in various ways. Donabedian defined the quality as "*the ability to achieve desirable objectives using legitimate means*" where the objective is almost always an achievable state of health. [7] In 1990, Institute of Medicine used the definition "*the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge*" which is arguably the most commonly used definitions of quality of care today.[8] According to World Health Organisation (WHO), healthcare should be safe (avoiding causing harm to the patients when providing care), effective (providing evidence-based care to the patients), efficient (using available resources efficiently), people-centered (patients' needs, preferences and values are taken into account when providing care), timely (providing care to patients in a timely manner, reducing waiting times and avoiding delay of care), equitable (providing care that does not vary between patients, for example, due to their socioeconomic status and ethnicity) and integrated (providing care that is coordinated across different levels of healthcare systems and providers). [1, 9, 10] (Figure 1)

Figure 1. Dimensions of quality of healthcare

To achieve high quality of care, WHO recently published a toolkit to improve the quality of care at national, sub-national, facility and community level. [1] Suggested steps in this toolkit included; defining a quality policy and strategy (for example, setting healthcare priorities, defining quality indicators and measures and alignment of information and data systems), defining quality standards (for examples, with established guidelines), regulation, measurement and evaluation of the quality, taking action by quality improvement interventions and engagement across different departments, sectors, stakeholders or other levels of the healthcare system. [1] Whichever level of the healthcare system is the level-of-interest, measurement and evaluation of the quality remains an essential step to ensure high quality of care and facilitate quality improvement initiatives. [1]

3.2. QUALITY MEASUREMENT

Quality needs to be measured continuously to track progress and to identify the room for improvement. [1] Measuring and quantifying the quality of care makes it possible to evaluate effects of quality improvement interventions and programmes and facilitate informed decision-making when managing health care. [1]

WHO summarized the quality measurement approach under five main steps, (1) definition of the framework (e.g., "what needs to be measured?"), (2) mapping quality indicators and sources (e.g., "which indicators are available and furthermore, are available indicators sufficient to measure what needs to be measured?"), (3) selection of quality indicators, (4) implementation (e.g., assessment of how the gathered information can be used) (5) assessment and visualisation of the data (e.g., "is the gathered information user-friendly?"). [1]

To quantify the quality of care, quality indicators have been used. Quality indicators provide comparative information to be used when monitoring, evaluating, providing feedback, benchmarking and managing healthcare services and for decision-making within and across healthcare systems. [10, 11]

There are different approaches to categorize quality indicators. [11] The commonly used Donabedian approach classifies quality indicators into three categories as structure, process and outcome indicators and is based on the assumption that good structure of care may lead to improved process of care, where improved process of care may lead to improved outcomes of quality of care. [12] (Figure 2)

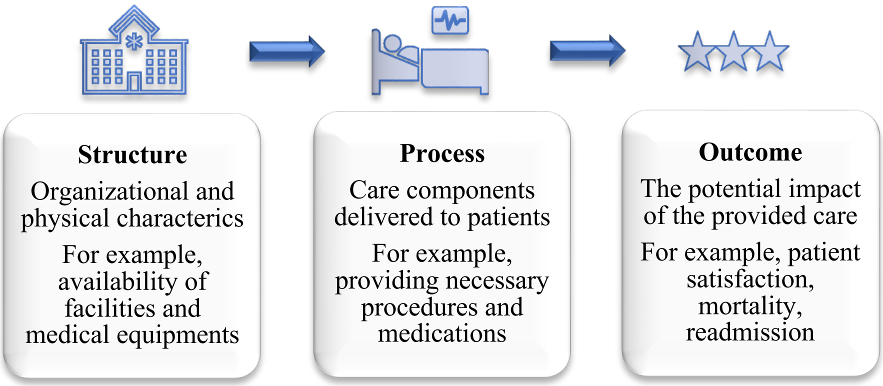

Figure 2. Donabedian model

Structure indicators focus on organisational and physical characteristics of where healthcare is provided. [2, 11] Number of beds in a hospital, existence of electronic patient record systems and adequacy of medical equipment in a healthcare provider are some examples for structural quality of care indicators.

Process indicators focus on actual care provided to patients, such as processes related to screening, diagnosis, and treatment of patients. [2, 11]

Outcome indicators aim to measure the impact of provided care on patients' status. [2, 11] Mortality, unplanned readmission, hospital-acquired infections, quality of life and patient satisfaction are some examples for outcome indicators.

While quality indicators provide valuable information, to achieve actionable results it is important that presented data and results are user-friendly, interpretable, and telling a relevant story for the audience. [2, 5] To translate the data obtained from the quality indicators into actionable information for the audience, it is important to consider who is the audience and what are their information needs. [5]

3.3. COMPOSITE MEASURES

A composite quality measure is a combination of multiple quality indicators into a single measure. [4]

3.3.1. RATIONALE FOR THE USE OF COMPOSITE MEASURES

Individual quality indicators carry important information that illustrate what has been done for patients and the quality of the provided care interventions. [11] While having this level of information is essential to monitor and improve the quality, the actionability of the provided information for the targeted audience should be considered. For example, using individual indicator level information to inform clinicians about their performance and their level of compliance with recommended processes in guidelines can be very useful. On the other hand, for example on the managerial level where the end user will assess quality of care in multiple healthcare providers by using multiple individual quality indicators, it can be a challenging and time-consuming task.

Some of the actors in a healthcare system and examples for their information needs and corresponding level of information for these needs are provided in Table 1.

Table 1. Examples for actors in a healthcare system and their information needs

Audience	Examples of potential questions related to assessment of quality of care	Information needs
Clinicians	"How can I improve the quality of care I provided to my patients?"	Individual indicator level
Patients	"Which healthcare provider is the best overall at delivering care for my condition?"	Composite measures
	"Which healthcare provider is the best at doctor-patient communication?"	Individual indicator level
Healthcare providers	"Is the overall quality of care improving in my hospital?"	Composite measures
	"Which components of the quality of care need to improve in my hospital?"	Individual indicator level
Regional level	"How well does our region perform compared to other regions?"	Composite measures
	"How can we improve the performance of the hospitals in this region? Where is the room for improvement?"	Individual indicators

National level	"Is the quality of care in our country improving over time?"	Composite measures
	"What is the impact of the current healthcare policies on the quality of care?"	Composite measures
International level	"How does our country perform compared to other high-income countries?"	Composite measures
Research	"Does overall adherence to clinical guidelines lead to better outcomes?"	Composite measures
	"Which aspects of quality have a stronger association with improved patient outcomes?"	Individual indicators

3.3.2. ADVANTAGES ASSOCIATED WITH COMPOSITE MEASURES

Using composite measures to evaluate the quality of care has some important advantages.

By definition, these measures summarize quality of care with a single number. This facilitates easier interpretation of the quality, easier communication with the public, tracking progress, assessment of overall effects of quality improvement programmes, guidelines or policies, benchmarking performance, comparison of healthcare providers, regions, countries or any level of interest and implementation of financial incentives that pay providers according to their overall performance. [3-5]

Reduced amount of data and condensed, comprehensive information. Using composite measures, the amount of data and information in quality reports that reader must process is decreased, leading to reduced information burden for the users. [3-5] At macro level where the end user will assess the quality of care in a large number of healthcare providers and is not directly interested in individual indicator achievement, composite measures can deliver an easier interpretation of overall quality.[3] For patients, an overall summary score can be more interpretable than information on multiple individual indicators. [3] Composite measures provide condensed and a more comprehensive assessment of quality and includes different dimensions of care in one measure. This makes it possible to track a broader range of quality metrics. [4]

Communication. Use of composite measures can be especially useful to inform the public and communicate complex and large amounts of data in an effective and simple way ensuring public and media attention. [5]

Easier comparisons. Comparison of healthcare providers, regions, countries, or any other level of interest can be a difficult task using multiple individual indicators as it would require assessment of a large number of scores for each the unit of comparison. Furthermore, it would require a similar trend in all individual indicators across unit of comparison as a higher achievement in some of the indicators and lower achievement in other indicators will not provide an easy and clear interpretation of which providers are doing the best. This concern is also valid for tracking the progress in healthcare providers, i.e., comparison of healthcare providers to their own performance from previous years. Using composite measures enable users to track the progress of care easier. [5]

Increased reliability. By combining multiple individual indicators, the number of observations on which the measurement is based will increase, which may lead to increased reliability. [3, 6, 13]

3.3.3. LIMITATIONS ASSOCIATED WITH COMPOSITE MEASURES

Despite important advantages of composite measures, these measures are also associated with important limitations.

Sensitivity and vulnerability of the methodology. It has been suggested that different methodologies to construct composite measures may lead to different results and different interpretations of quality.[6, 14-16] While different construction methods can be also an advantage of composite measures (for example, construction of a composite measure can be modified according to a specific audience and their needs and definitions of "high quality"), the sensitivity of composite measures to the used method may lead to unreliable and misleading conclusions.

Misleading conclusions. If composite measures are not constructed in a reliable, robust and sound way, obtained results and conclusions may be misleading. [5] For example, if a composite measure is based on easily obtainable but not important indicators and important but hard to measure aspects of quality are ignored, the results derived from these composite measures may lead to inappropriate decisions.

Susceptibility to misuse. If the construction process of the composite measures is not transparent and not justified, these measures can be susceptible to misuse. For example, by adjusting weights of individual indicators, a desired score or rank may be achieved even though these scores or rankings may not accurately reflect the provided quality of care. [17]

Masking important information. Combining multiple individual indicators into one single measure has the inherent problem of masking important information regarding potentially important aspects of quality of care. [4] Important information may be lost, and use of composite measures may lead to too simplistic interpretations. [4]

3.3.4. CONSTRUCTION METHODS

The quality and soundness of a composite measure is highly dependent on the quality of the underlying data, a carefully defined framework and a methodologically sound construction process.[5]

While there is not a standard established method to construct composite measures, some steps are suggested to take into account when constructing these measures.[5] Some important steps of building composite measures and some examples of considerations for each step are illustrated in Figure 3.

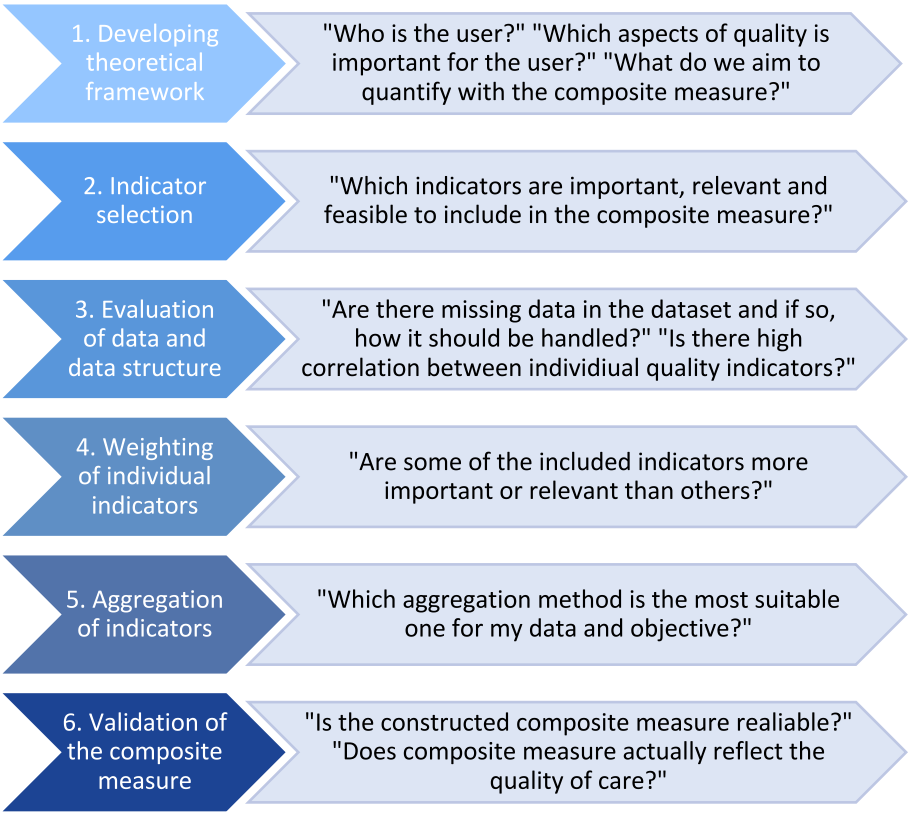

Figure 3. Construction steps for composite measures

The construction of composite indicators starts with a defined framework.[5] A theoretical framework should define (1) “why”, why do we construct this composite measure and what we aim to measure with this composite measure exactly, (2) “for who”, who is the user of the composite measure to be created as different users may have different information needs and different values and definitions of high quality, (3) “which”, which aspects of quality are important and which indicators can be relevant for defined audience and the defined objective, leading to the potential inclusion and exclusion criteria for the quality indicators to be considered. [4, 5]

After defining the aim and potential inclusion criteria for the quality indicators, decision of which indicators are useful and relevant to fulfil the defined objective should be made.[5] Reliability and usefulness of composite indicators are linked to underlying individual indicators. [5] Some aspects to consider when selecting quality indicators can be their relevance, feasibility, importance, discriminatory power, timeliness, availability, accuracy and completeness.[5]

Next steps include investigation of dataset and data structure. If there is missing data, imputation of missing data and potential approaches for imputation and advantages and limitations of different approaches can be considered. [5] If individual indicators have different measurement scales, normalisation of data is required prior to aggregation. [5] Analysing the underlying data structure and interrelations between individual indicators should be considered. For example, if a composite measure is constructed with indicators which are highly correlated with each other and potentially measuring the same aspect of quality, this may lead to “*indicator rich but information poor*” composite indicators. [5]

Weighting of individual indicators before aggregation enables users to differentiate between individual indicators and to emphasize certain aspects of quality which may be more important or relevant (or other defined criteria) to achieve the outcome of interest. Some of the weighting approaches used in the literature are explained below.

Weighting individual indicators equally usually implies that there are no grounds to differentiate between quality indicators and they are equally relevant and important to measure the overall quality of care. [5] It should be considered that, even if the equal weights are assigned to each indicator with the intention of “equal importance” of each aspect, it may result differently. For example, if there are multiple indicators measuring the same aspects of quality, even though we assign equal weights to each indicator, this component of quality will be measured multiple times and will weigh more in the composite indicator. This issue is also referred as “double weighting”. [5] If there are large differences in number of observations for individual indicators, this may also create unbalance, even if each aspect of quality weighted equally. [5] This shows the importance of exploring the data and data structure prior to combining quality indicators.

If there is high correlation between individual indicators, this can imply that the indicators share similar underlying traits, measure similar aspects of the quality of care and there may be overlapping information in these indicators. In such cases, weights obtained from principal component analysis (PCA) or factor analysis (FA) can be considered. PCA or FA weights may be helpful to avoid assigning higher weights unintentionally to one aspect of care. [5]

In regression weights approach, weights can be assigned to each individual indicator according to their size of association with the desired outcome. Using regression weights approach, indicators which have a greater association with the outcome will receive higher weights. This method can be preferred when there is a "gold-standard end point".[16]

After weighting each component of quality, these indicators can be aggregated using different schemes.

Opportunity-based composite scores (OBCS) can be investigated under two sub-categories: overall percentage and patient average.

Overall percentage composite scores are obtained as the total number of care components provided to all patients in the hospital, divided by the total number of care components that were relevant for those patients' care, indicating what percentage of care patients "received" that they "should" have received in a hospital. [6, 18]

Using overall percentage approach, composite score for a hospital can be calculated as;

$$\text{Overall percentage score} = \frac{\sum_{i=1}^{n} \text{Total number of care indicators delivered to patient}_i}{\sum_{i=1}^{n} \text{Total number of care indicators relevant to patient}_i\text{'s care}}$$

where n denotes the number of patients in the hospital.

In patient average method, each patient receives a score according to fulfilment of care indicators that are relevant to them. [6, 18] While patient average composite scores are calculated similarly to overall percentage (provided care divided by relevant care), in patient average method scores are calculated on patient-level and then can be averaged to obtain hospital or other level-of-interest composite scores. Overall percentage composite scores are calculated directly on provider level.

Using patient average approach, composite score for a hospital can be calculated as,

$$\text{Score for patient}_i = \frac{\text{Total number of care indicators delivered to patient}_i}{\text{Total number of care indicators relevant to patient}_i\text{'s care}}$$

$$\text{Patient average score for hospital} = \frac{\sum_{i=1}^{n} \text{Score for patient}_i}{n}$$

where n denotes total number of patients in the hospital.

All-or-none (AON) scoring is the strictest method to construct composite indicators. In AON, only complete care of patients is rewarded. Each patient receives either 1 (patient received all of the care components that the patient was eligible for) or 0 (at least one of the necessary care components was not provided to patient). [3, 6, 18] Patient scores can be averaged to obtain hospital-level composite scores.

Using AON, composite indicators can be calculated as

$$\text{AON score for each patient} = \begin{cases} 1, & \text{All relevant care components are provided to the patient} \\ 0, & \text{At least one relevant care component is not provided to the patient} \end{cases}$$

$$\text{Hospital-level AON score} = \frac{\sum_{i=1}^{n} \text{Score for } patient_i}{n}$$

where n denotes the number of patients in the hospital.

In 70% standard composite scoring approach, composite scores are calculated in a similar way to AON, with a lower threshold. Each patient receives either 1 (at least 70% of eligible care components were provided to the patient) or 0 (patient received less than 70% of eligible care components). [6, 18]

Using 70% standard approach, composite indicators can be calculated as

$$\text{70\% standard score for each patient} = \begin{cases} 1, & \text{At least 70\% of the eligible care components were delivered to the patient} \\ 0, & \text{Patient received less than 70\% of necessary care components} \end{cases}$$

$$\text{Hospital-level composite score} = \frac{\sum_{i=1}^{n} \text{Score for } patient_i}{n}$$

where n denotes the total number of patients in the hospital.

While these are some of the most common approaches to construct composite measures, numerous other approaches exist. For example, patient/survey weights (weights are assigned according to patients' judgement of what is more important for them), benefit of doubt weights (assigning weights in a way to maximize healthcare providers' score), Bayesian weights and item response theory weights.

Next and the final step in construction process of composite measures is validation. Even if all the individual indicators used in the composite measures are already validated measures of quality, it is still recommended to validate the final composite measure. [19]

While validity relates to accuracy of the measurement, reliability relates to the consistency of the measure.[20] Defining the framework (intended aim and intended user) in the beginning of the construction process is important to ensure validity and reliability of a composite measure.[5]

Some of the validation types are content validation (a subjective judgement regarding whether all necessary care aspects are included in the composite measure), construct validation (whether composite indicator measures what is aimed to be measured), face validity (a subjective judgement and agreement by an expert panel or targeted stakeholders that the composite measure is useful and valid), criterion validity (the extent to which the constructed composite measure correlates or predicts the gold-standard end point). [4]

After step-by-step construction process is completed, the most appropriate way to communicate this information obtained by composite measures, including visualisation methods for the targeted audience, can be considered. [5]

3.4. THE DANISH HEALTHCARE SYSTEM AND QUALITY MEASUREMENT IN THE DANISH CONTEXT

Denmark has public healthcare system where the access to the majority of healthcare services is free of charge for all Danish residents and mainly financed by taxes. [21-24]

In Denmark, healthcare is organized on three administrative levels: national, regional and municipal level. [21-23] While state is in charge of legal frameworks, regulation, supervision, and fiscal functions, regions are in charge of hospitals and self-employed general practitioners, and municipalities are responsible for health promotion and disease prevention. [23]

As also appraised in a report by Organisation for Economic Co-operation and Development (OECD), Denmark has impressive quality assurance and improvement mechanisms in place that includes quality improvement projects, establishment of quality standards and guidelines and extensive and well-developed clinical registries. [24]

One of the important initiatives in Denmark towards quality improvement was the Danish National Indicator Project. [25] The Danish National Indicator project was established in 2000 with an aim to improve quality of care delivered to the patients in six specific disease areas and participation was mandatory for healthcare providers.[25] Quality indicators and standards were developed.[25] The performance of each healthcare provider was monitored and assessed, and the results regarding their performance on achievement of quality indicators and standards were shared with respective healthcare providers. [25] Furthermore, national and regional audits were conducted to discuss and analyse these results and facilitate potential improvement processes. [25] Results from this project revealed significant improvement in the delivery of quality of care. For example, for stroke care, the proportion of patients with timely admission to a stroke unit increased from 77% in 2003 to 91% in 2008. [26] Similar trends were observed in delivery of all process indicators to stroke patients, whereas 30-day mortality was declined from 12% in 2003 to 10% in 2008. [26] These results showed the potential effects of measurement and documentation of quality, establishing quality standards and providing continuous feedback on improvement of quality of care.

In 2015, the National Quality Programme was introduced. The national objectives for the Danish healthcare system were defined, which included improved patient pathways, better quality of care with a special focus for chronically ill and elderly patients, increased survival rates, healthy life years and patient safety, timely diagnosis and treatment, greater patient involvement and a more effective healthcare system. [21, 27]

Today, Denmark has a total of 85 well-developed clinical quality registries, managed by The Danish Clinical Quality Program– National Clinical Registries (RKKP). [28] These registries contain patient-level information used for evaluation and improvement of quality of care. [28] The quality indicators in these clinical quality registries principally reflect the recommendations from established national clinical guidelines. [22, 25] These registries overall have high coverage, as this quality improvement programme is exempt from patient consent and it is mandatory for all Danish hospitals to register patients in corresponding registries. [28] Furthermore, in Denmark each resident receives a 10-digit identification number either at birth or immigration that facilitates the linkages among different Danish registries. [29] Hence, these registries are very valuable resources for research.

3.4.1. HEART FAILURE AND MONITORING THE QUALITY OF HEART FAILURE CARE IN DENMARK

European Society of Cardiology defines heart failure (HF) as "*a clinical syndrome characterized by typical symptoms (e.g. breathlessness, ankle swelling and fatigue) that may be accompanied by signs (e.g. elevated jugular venous pressure, pulmonary crackles and peripheral oedema) caused by a structural and/or functional cardiac abnormality, resulting in a reduced cardiac output and/or elevated intracardiac pressures at rest or during stress.*".[30] HF creates heavy burden for patients, their families and society. [31-33] Due to HF's worldwide high prevalence, poor prognosis and heavy burden, measurement and improvement of quality of HF care is critical.[34, 35] To evaluate and improve the quality of care provided to patients with HF, many countries have established clinical registries. [36-39]

The Danish Heart Failure Registry (DHFR) was established to monitor and assure the delivery of high quality of care to the patients with HF in Denmark. [39] Since its establishment in 2003, the achievement rates of process indicators are substantially increased, while 1-year mortality for patients with HF declined. [40] To be registered in the DHFR, the diagnosis must be verified by a cardiologist.[39] The DHFR contains information on processes and outcomes of quality of care, as well as prognostic factors. [39]

3.4.2. ISCHEMIC STROKE AND MONITORING THE QUALITY OF ISCHEMIC STROKE CARE IN DENMARK

WHO defines stroke as *"rapidly developed clinical signs of focal (or global) disturbance of cerebral function, lasting more than 24 hours or leading to death, with no apparent cause other than of vascular origin".*[41] Stroke is a major cause of mortality and disability worldwide and while mortality rates for stroke is decreasing, the overall burden of the stroke remains high.[42, 43] Previous studies revealed that timely delivery of high quality of care to ischemic stroke patients is associated with improved patient outcomes, including lower mortality. [44-47] To monitor, track and evaluate the quality of care delivered to the stroke patients, many countries have established national clinical registries.[48]

To monitor and improve the quality of care delivered to the patients with stroke in Denmark, the Danish Stroke Registry (DSR) was established in 2003.[49] The DSR consists of prospectively collected individual-level data for all adult patients (aged 18 or older) admitted to Danish hospitals with acute stroke. In the DSR, stroke is defined according to WHO criteria provided above. The DSR includes individual-level information on processes and outcomes of care as well as information on prognostic factors. [49] The sensitivity and predictive value of the registration of patients with stroke in the DSR was estimated to be over 90%. [50]

CHAPTER 4. PAPER I

This chapter is based on the paper "*Composite measures of quality of health care: Evidence mapping of methodology and reporting*" by Kara P, Valentin JB, Mainz J, Johnsen SP that was published in Plos One and attached as appended paper I. [51]

4.1. AIMS

The main aims of this study were as follows,

i. To identify the methods that were used in the literature to construct composite measures to quantify quality of care based on process of care indicators.
ii. To investigate methodological considerations behind these methods: (1) if authors provided any justification regarding their choice of methodology to build composite measures, (2) if authors provided any information on advantages or limitations of using composite measures rather than using individual quality indicators and (3) if authors informed readers regarding the presence of other approaches to construct composite measures.

4.2. METHODS

A literature search was conducted on PubMed and EMBASE databases to identify the publications that used composite indicators to quantify the quality of care. Search string is available in appended paper I.

We excluded studies if (1) indicators other than process indicators were used in the composite measures, (2) indicators related to access or utilization were used, (3) indicators were patient-reported, (4) the type of publication was protocol, review, editorial, comment, letter, trial design or purely methodological, (5) publication language was not English and (6) full-text was not available. Detailed exclusion reasons and PRISMA diagram was provided in appended paper I.

For included studies, each paper was assessed for (1) construction methodology for composite measures and (2) whether methodological considerations regarding composite measures were addressed. [51]

4.3. RESULTS

Out of 2711 identified publications, 145 publications satisfied the full inclusion criteria and consequently, included in the study. References for included studies are provided in the appended paper I.

Included papers were categorized in two categories according to their main aim; to support operational use or for research purposes. Further explanation and examples were provided in appended paper I.

For aggregation, majority of the included studies constructed composite indicators using OBCS approach (89/145), whereas AON scoring was the second most used approach (48/145) to construct composite measures.

For weighting of each indicator before aggregation, majority of the publications relied on equal weights. Other approaches were limited and included: expert weights (7/145), PCA-based weights (3/145), item response theory weights (3/145), Bayesian weights (1/145), benefit of doubt weights (1/145) and regression weights (1/145).

For methodological considerations addressed in the publications in relation to use of composite indicators, any rationale regarding the choice of methodology was provided in 36 publications. Out of 145 publications, 42 studies provided any advantage of using composite indicators and 22 publications mentioned any limitation of using composite indicators. [51] (Table 2)

Table 2. Methodological information in the reviewed publications [51]

Methodological information	**Number of papers (%)**
Communication of the construction method: Any rationale regarding the choice of methodology was addressed/referenced	36 %25%)
Communication of limitations: Any limitation associated with composite measures was addressed/referenced	22 (15%)
Communication of advantages: Any advantage regarding use of composite measures was addressed/referenced	42 (29%)
Informing the reader regarding the presence of other approaches, only for papers that used a single approach.	10 (8%)

Detailed information regarding construction method and methodological considerations in each included publication are available in the appended paper I.

4.4. DISCUSSION

4.4.1. MAIN RESULTS

Main findings of this study revealed that even though numerous approaches to construct composite measures exist, the variation in the construction process of composite measures to measure process quality of health care was modest. Furthermore, although using different approaches may lead to different results and consequently facilitate unreliable interpretations of quality, the majority of the included publications did not inform the reader about the rationale for selecting preferred methodology, limitations associated with composite measures and presence of other construction methodologies. [51]

4.4.2. COMPARISON WITH THE EXISTING LITERATURE

To our knowledge, this was the first study that systematically identified and examined studies that used composite measures to quantify the quality of care based on process indicators. [51]

4.4.3. METHODOLOGICAL CONSIDERATIONS

In study I, we recognized certain limitations. First, this review was restricted to peer-reviewed publications indexed in the PubMed and EMBASE databases. Second, we only included publications that are in English, consequently relevant publications in other languages might have been excluded. Lastly, although recommended in guidelines, we did not check whether validation of the composite measures was performed in the included studies, as this broad topic was beyond the scope of the study. [51]

4.5. CONCLUSION

The variation in the existing studies regarding construction of composite measures to measure process quality of health care was modest and the attention to methodological considerations was limited and/or not communicated with the reader clearly.

CHAPTER 5. PAPER II AND PAPER III

This chapter is based on the paper II: "Composite measures of quality of care for patients with heart failure in Denmark: A nationwide cohort study" which is in preparation stage and paper III: "Composite measures of quality of care for patients with ischemic stroke in Denmark: A nationwide cohort study" which is in preparation stage. [52, 53]

Paper II and Paper III are attached to this thesis as appended paper II and appended paper III, respectively.

5.1. AIMS

5.1.1. AIMS: PAPER II

In study II, we constructed multiple composite measures of quality of incident HF care based on process indicators selected by an expert panel and then combined using multiple weighting and aggregation methods. Using these composite measures, we aimed,

I. To rank hospitals using different types of composite measures based on various weighting and aggregation schemes and to investigate how hospitals' rankings change according to the method that was used to construct composite measures.

II. To investigate the association between different types of composite quality measures and 1-year all-cause mortality for patients with incident HF.

5.1.2. AIMS: PAPER III

In study III, we constructed multiple composite measures of quality of incident ischemic stroke care based on quality indicators selected by an expert panel and then combined using multiple weighting and aggregation methods. Using these composite measures, we aimed,

I. To rank hospitals using different types of composite measures based on various weighting and aggregation schemes and to investigate how hospitals' rankings change according to the method that was used to construct composite measures.

II. To investigate the association between different types of composite quality measures and 30-day all-cause mortality for patients with incident ischemic stroke.

III. To investigate predictive accuracy of different types of composite measures of quality of care in relation to 30-day all-cause mortality.

5.2. METHODS

Study II and Study III were nationwide cohort studies conducted based on data from registries in Denmark. Using 10-digit unique civil registration numbers assigned to all Danish residents, the linkages between different Danish registries were created.

5.2.1. METHODS: PAPER II

Data Sources
The data sources used in study II were The Danish Heart Failure Registry (DHFR), The Danish National Patient Registry (DNPR), The Danish National Prescription Registry (DPR), The Danish Civil Registration System (CRS) and Statistics Denmark.

The DHFR provided the source population, information on quality indicators, lifestyle factors, clinical characteristics, and history of hypertension. The DNPR and DPR were used to identify patients with comorbidities. Statistics Denmark provided information on annual family income and highest level of completed education. The CRS provided information on sex, date of birth, vital status, and emigration status.

Further information on used data sources for each covariate were provided in appended paper II.

Study Population
Using DHFR, we identified all adult patients diagnosed with incident HF from 01.01.2008 till 31.12.2018. We excluded patients if their vital status was missing, if they resided in Denmark less than 5 years upon hospital contact, if they were not eligible for any of the quality indicators, if they died on or before the index date, if they were treated in hospitals with less than 100 patients in total for defined study period, if the index date for composite measure was after the end of follow-up and if their index date was after study period, 31 December 2018.

Index date was defined as the first day after the time window to provide process indicators according to clinical guidelines expires. Information on index date for each process indicator and composite indicator can be found on appended paper II.

Expert panel
We conducted an expert panel to (1) select clinically meaningful process indicators to use in our study and (2) to weigh selected indicators to construct composite indicators with expert weights.

Expert panel consisted of three experts from DHFR steering group with expertise in both DHFR and clinical work with HF care. The entire expert panel process is provided in appended paper II.

Covariates

In study II, we included age, sex, high alcohol consumption, smoking, left ventricular ejection fraction (LVEF), New York Heart Association (NYHA) classification, history of comorbidities (acute myocardial infarction, ischemic stroke, diabetes mellitus, chronic obstructive pulmonary disease, hypertension, chronic kidney disease, obesity, valvular heart disease and obstructive sleep apnea), annual family income and highest level of completed education as covariates. Covariate definitions and the used data sources for covariates are provided in appended paper II.

Statistical Analysis

Overall quality of HF care for each hospital were obtained using multiple methods based on different weighting and aggregation schemes: overall percentage with equal weights, patient average with expert weights, patient average with regression weights, AON scoring and 70% standard approaches, with unadjusted and multivariable adjusted analyses.

For patient average methods where each patient received a score between 0 and 1, we used fractional probit regression, for AON scoring and 70% standard approaches where the patient scores were dichotomous, we used logistic regression and for overall percentage with equal weights, we used linear regression and composite scores were obtained as marginal proportions. For patient average with regression weights approach, we used Cox regression to assess each individual indicator's association with 1-year all-cause mortality.

Calculation of composite indicators and descriptions for methods are explained further in appended paper II.

Ranking of healthcare providers

Unadjusted and multivariable-adjusted hospital rankings were obtained according to hospitals' composite scores based on different methods.

Rankability

We drew bootstrap samples with 1,000 replications from the original dataset. Using one of the construction methods (patient average with expert weights 1) each hospital received 1,000 rankings based on bootstrap samples and bootstrap obtained 95% CIs were provided. This was used as an indication whether the variation in hospital rankings based on different methods was larger than random variation within one methodology.

Association with 1-year all-cause mortality
To assess the association between composite measures of HF care and 1-year all-cause mortality, we performed Cox proportional hazards regression analysis. For patients with incident HF, the time at risk started on the first day after the latest possible day to provide necessary care to patients in order to comply with clinical recommendations (85 days after first hospital contact). The time at risk ended on the date of death, loss to follow-up or at the end of 1-year follow-up, whichever came first. For each composite scoring method, unadjusted and multivariable-adjusted hazard ratios and 95% CIs were calculated.

Sensitivity analysis
We conducted a sensitivity analysis where the time at risk started at the day of first hospital contact. [52]

We conducted all statistical analyses using Stata 16 (StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC.).

5.2.2. METHODS: PAPER III

Data sources
The data sources used in study III were The Danish Stroke Registry (DSR), the DNPR, the DPR, the CRS and Statistics Denmark.

The DSR provided information on lifestyle factors, sociodemographic factors, severity of stroke, comorbidities and indicators related to processes and outcomes of care. The DNPR and the DPR were used as supplementary data sources to further identify patients' comorbidities. Statistics Denmark provided information on annual family income, highest completed education and emigration. The CRS provided individual-level information on sex, date of birth and vital status.

Detailed information on used data sources were provided in appended paper III.

Study population
Using DSR, we identified patients with incident acute ischemic stroke from 01.01.2016 till 31.12.2018. We excluded patients if information regarding vital status was missing, if the time interval between symptom onset and hospital admission was more than one week, if they died within two days of hospital admission, if they lived in Denmark less than 5 years upon hospital admission or moved out of Denmark during follow-up period, if they were treated in hospitals with less than 100 ischemic stroke patients in total for the defined study period and if they had missing information on selected quality indicators for the study.

Expert panel
We conducted an expert panel to (1) select clinically meaningful process indicators to use in our study and (2) to weigh selected indicators to construct composite indicators with expert weights.

Expert panel consisted of three experts from the DSR steering group with expertise in both the DSR and clinical work with ischemic stroke care. The entire expert panel process is provided in appended paper III.

Covariates
Included covariates in this study were age, sex, type of residence, cohabitation status, high alcohol consumption, smoking, stroke severity (Scandinavian Stroke Scale, SSS), history of comorbidities (acute myocardial infarction, diabetes mellitus, hypertension, atrial fibrillation, and intermittent claudication), annual disposable family income and highest level of completed education.

Statistical Analysis
Overall quality of ischemic stroke care for each hospital were obtained using multiple methods based on different weighting and aggregation schemes: overall percentage with equal weights, patient average with expert weights, patient average with regression weights, AON scoring and 70% standard approaches, with unadjusted and multivariable adjusted analyses.

Composite measures were obtained in a similarly to study II, described above. In study III, for the patient average with regression weights approach, we used logistic regression to assess each individual indicator's association with 30-day all-cause mortality.

Calculation of composite indicators and descriptions for methods are explained further in appended paper III.

Ranking of healthcare providers
Unadjusted and multivariable-adjusted hospital rankings were obtained according to hospitals' composite scores based on different methods.

Rankability
Rankability of composite measures was assessed similarly to study II, described above.

Association with 30-day all-cause mortality
We investigated the association between patient-level composite measures and 30-day all-cause mortality, using logistic regression, with unadjusted and multivariable adjusted analysis. Odds ratios and 95% CIs were obtained.

Predictive accuracy of composite measures
We examined the predictive accuracy of composite measures in predicting 30-day mortality, performing apparent validation and 5-fold cross validation. The incremental predictive value of the composite measures was assessed.

Sensitivity analysis
We performed a sensitivity analysis using a different index date where the risk for the patients started 14 days after hospital admission. [53]

We conducted all statistical analyses using Stata 16 (StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC.).

5.2.3. ETHICS

The Danish Health Data Authority, the Danish Data Protection Agency and Statistics Denmark approved study II and study III. Registry-based studies do not require approval from an ethics committee according to Danish law.

5.3. RESULTS

5.3.1. RESULTS: PAPER II

Study population
From 01.01.2008 till 31.12.2018, we identified 41,968 patients with incident HF. After excluding patients according to the exclusion criteria given under methods section, the final study population included 30,739 patients. The exclusion process and patient characteristics are provided in the appended paper II.

Results from expert panel
Expert panel selected five indicators out of seven indicators that were available in the DHFR. Selected indicators to be included in the composite scores are provided in Table 3.

Table 3. Selected indicators for composite measures of quality of HF care [52]

Process indicator	Indicator definition
Treatment with ACE inhibitor/ARB	Started, attempted to start or already on ACE inhibitor/ARB, within 8 weeks after first hospital contact, for patients with LVEF<=40%
Treatment with beta-blockers	Started, attempted to start or already on beta-blockers, within 12 weeks after first hospital contact, for patients with LVEF<=40%
Treatment with MRA	Started, attempted to start or already on mineralocorticoid receptor antagonists, within 12 weeks after first hospital contact, for patients with LVEF<=35%
Exercise training	Referral for exercise training by a physiotherapist in hospital, within 12 weeks after first hospital contact, for patients with LVEF<=40%
Patient education	Start-up of structured individualized patient education, within 12 weeks after first hospital contact, for patients with LVEF<=40%

We used three set of expert panel weights. Expert weights 1 (EW1) indicated each indicator was equally important (equal weights), whereas in expert weights 2 (EW2) indicators related to medication weighted more and in expert weights 3 (EW3), patient education received the highest weight. Weights of each indicator for each set of expert weights are provided in appended paper II.

Rankings of hospitals

Hospital rankings based on different composite scoring methodologies provided above were obtained and plotted, with and without adjustment. (Figure 4, Figure 5)

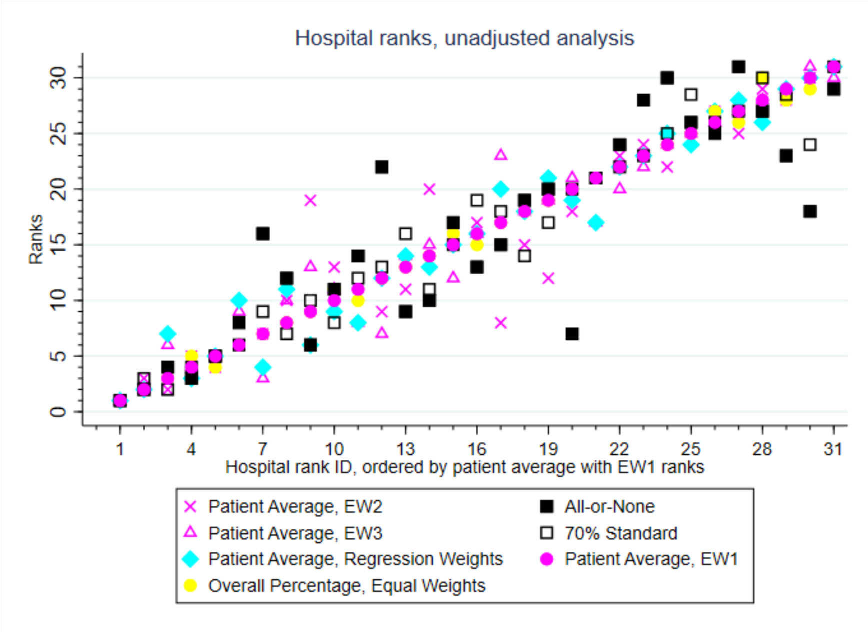

Figure 4. Unadjusted hospital rankings [52]

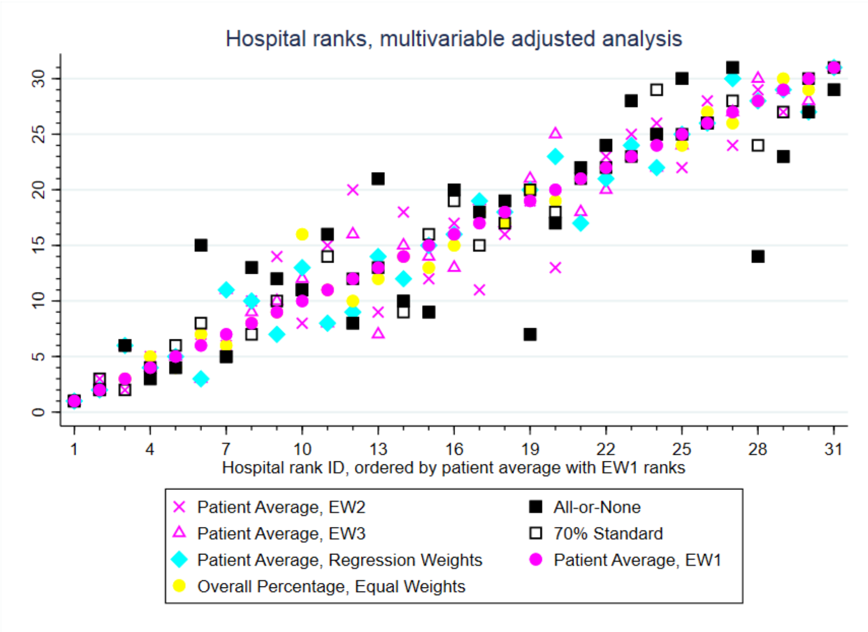

Figure 5. Multivariable-adjusted hospital rankings [52]

Both for unadjusted and adjusted analysis, majority of hospitals' rankings differed according to composite scoring methodology. (Figure 4, Figure 5)

Rankability

Figure 6 and Figure 7 showed that the variation between hospital rankings based on different methods was larger than random variation for many hospitals, which cannot be solely explained by chance, both for unadjusted and adjusted analysis.

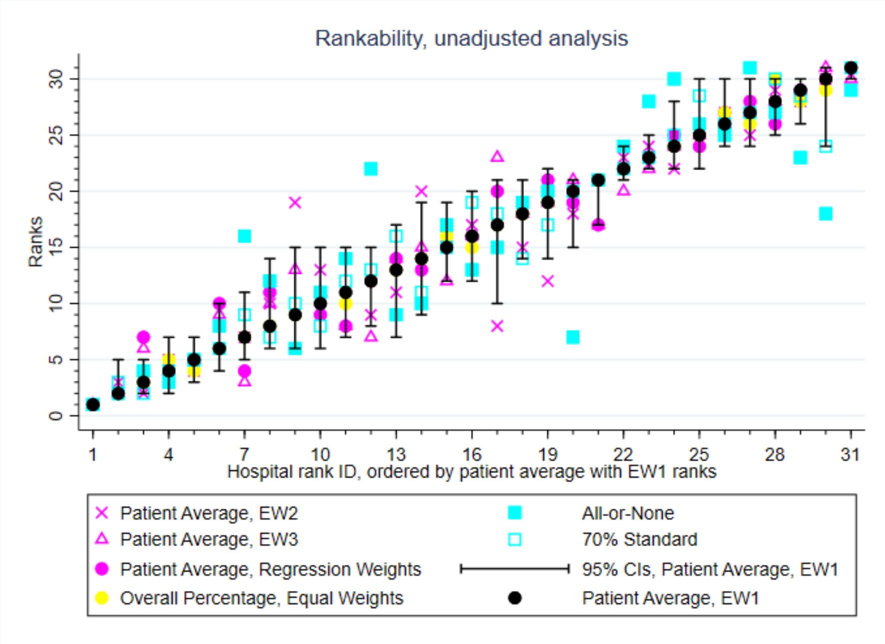

Figure 6. Rankability, unadjusted analysis [52]

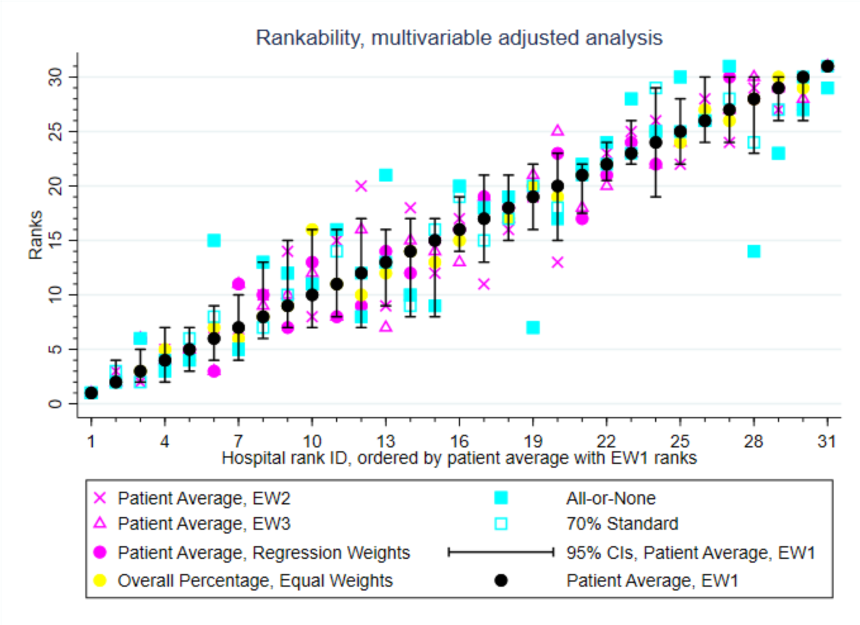

Figure 7. Rankability, multivariable adjusted analysis [52]

Association with 1-year all-cause mortality

The association between composite measures of quality of HF care and 1-year all-cause mortality are provided in Table 4.

Table 4. Association between composite measures of quality of HF care and 1-year mortality [52]

Composite measures	Hazard ratios (95% CIs)	
	Unadjusted	Multivariable adjusted
Patient average, EW1	0.979 (0.978-0.981)	0.988 (0.986-0.990)
Patient average, EW2	0.982 (0.980-0.983)	0.990 (0.988-0.991)
Patient average, EW3	0.979 (0.978-0.981)	0.988 (0.986-0.989)
Patient average, Regression weights	0.979 (0.977-0.980)	0.987 (0.986-0.989)
AON	0.430 (0.377-0.491)	0.631 (0.551-0.723)
70% standard	0.464 (0.430-0.501)	0.643 (0.593-0.698)

The results for patient average methods show the effect of one percent change in composite score on 1-year all-cause mortality whereas the AON and 70% standard methods show the hazard ratio of one group (complete care and at least 70% of required care, respectively) compared to the other group (not complete care and less than 70% of necessary care, respectively). Therefore, it is important to emphasize the magnitude of association obtained by patient average methods are not directly comparable to results obtained by AON and 70% standard methods.

Results showed that improvement in all types of composite scores were associated with lower hazard of experiencing 1-year mortality. For example, one percent increase in patient average with EW3 score reduces the hazard of experiencing the event of interest, 1-year mortality, by 2.1%. for the unadjusted analysis. For further interpretations, please see appended paper II. [52]

Sensitivity analysis

Results from the sensitivity analysis showed that hazard ratios were slightly lower for patient average and 70% standard methods and higher for the AON method, both for adjusted and unadjusted analysis. For more detailed information, please see appended paper II. [52]

5.3.2. RESULTS: PAPER III

Study population

From 01.01.2016 till 31.12.2018, we identified 26,297 patients with incident ischemic stroke. After excluding patients according to the exclusion criteria given under methods section, the final study population included 24,324 patients. The exclusion process and patient characteristics are provided in the appended paper III.

Results from expert panel

According to the results from the expert panel, the composite measures included 11 process indicators of quality of ischemic stroke care and are provided in Table 5.

Weights for each process indicator assigned by the expert panel are provided in the appended paper III.

Table 5. Selected indicators for composite quality measures of ischemic stroke care [53]

Quality indicators	**Definition**
Thrombolysis	Door to needle time ≤1 hour, for the patients who were treated with thrombolysis.
Timely admission	Admission to a stroke unit within the first day of hospital admission.
Platelet inhibitory treatment	Providing platelet inhibitor therapy within the second day of admission, for patients without atrial fibrillation.
Oral anticoagulation therapy	Providing oral anticoagulation therapy within 14 days after admission, for patients with atrial fibrillation.
CT / MRI scan	Performing CT/MR scan within 6 hours after hospital admission.
Assessment by a physiotherapist	Assessment by a physiotherapist regarding need of rehabilitation within the second day of admission.
Assessment by an occupational therapist	Assessment by an occupational therapist regarding need of rehabilitation within the second day of admission.
Early mobilization	Mobilization of patients on the day of admission.
Indirect swallowing test	Performing indirect swallowing test on the day of admission.
Ultrasound CT / MRI angiography of the carotid artery	Performing ultrasound CT/MR-angiography of carotid arteries within the fourth day of admission.
Timely carotid endarterectomy	Performing carotid endarterectomy within 14 days after admission, for the patients eligible for the procedure.

Rankings of hospitals

Hospital rankings based on different type of composite scoring methodologies were obtained and plotted, with and without adjustment. (Figure 8, Figure 9)

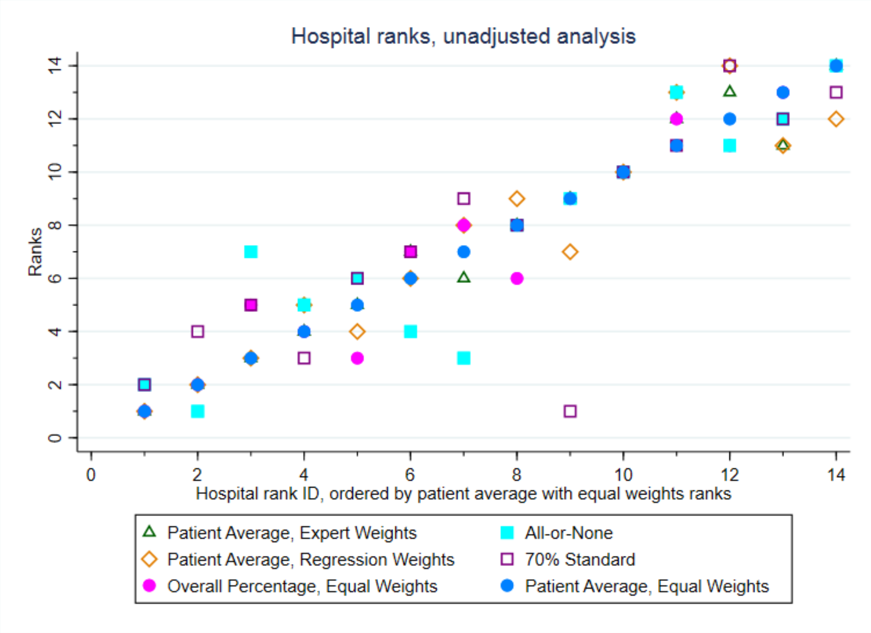

Figure 8. Unadjusted hospital rankings [53]

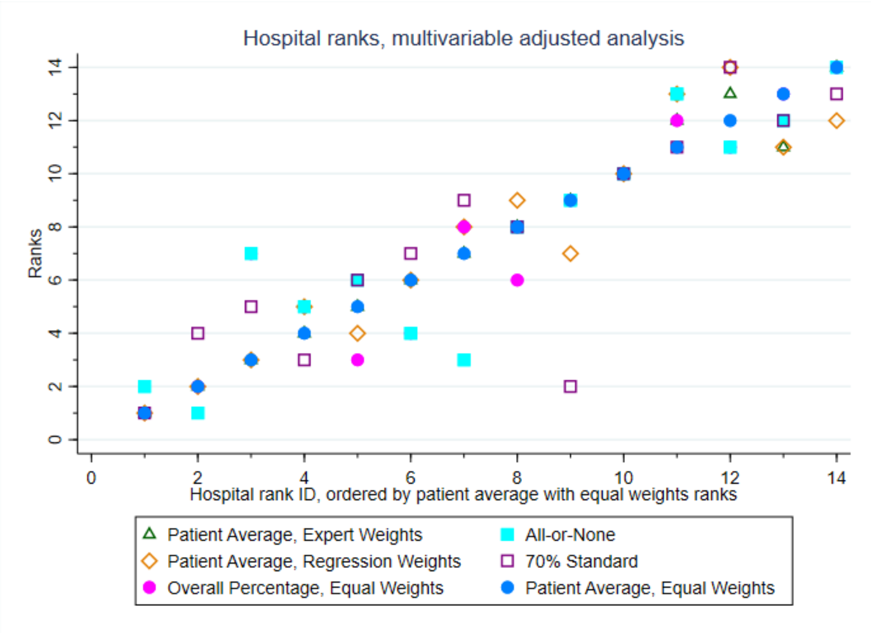

Figure 9. Multivariable adjusted hospital rankings [53]

Rankability

Figure 10 and Figure 11 revealed that the variation between rankings based on different methods were larger than random variation for some of the hospitals and this cannot be solely explained by chance, both for unadjusted and adjusted analysis.

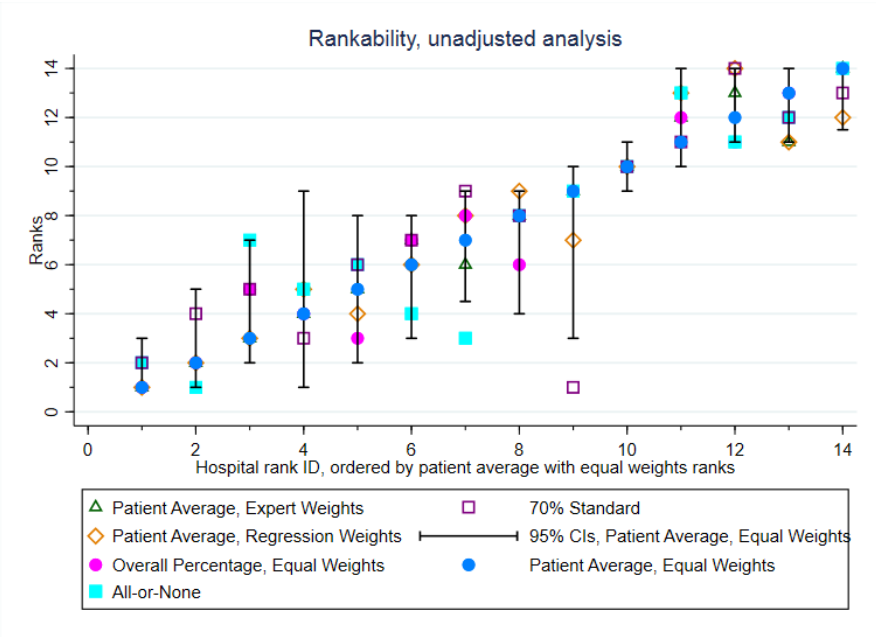

Figure 10. Rankability, unadjusted analysis [53]

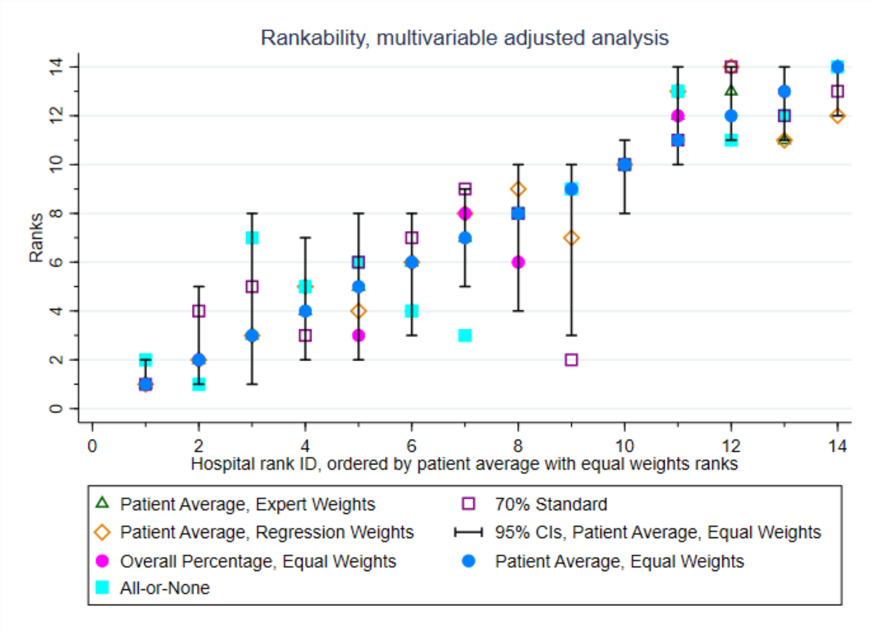

Figure 11. Rankability, multivariable adjusted analysis [53]

Association between composite measures and 30-day all-cause mortality
Improvement in all types of composite measures, except all-or-none scoring, was significantly associated with lower 30-day all-cause mortality. (Table 6)

Table 6. Association between composite measures of quality of ischemic stroke care and 30-day mortality [53]

Composite measures	**Odds ratios (95% CIs)**	
	Unadjusted	**Multivariable adjusted**
Patient average, Equal weights	0.993 (0.991-0.995)	0.995 (0.993-0.998)
Patient average, Expert weights	0.993 (0.991-0.995)	0.995 (0.992-0.997)
Patient average, Regression weights	0.992 (0.989-0.994)	0.995 (0.993-0.997)
AON scoring	0.960 (0.867-1.063)	1.084 (0.966-1.218)
70% standard	0.596 (0.525-0.677)	0.625 (0.540-0.723)

For patient average composite scores, odds ratios present the effect of one percent change in composite quality score on outcome, for example, one percent increase in patient average with regression weights is associated with 0.8% lower mortality, whereas for the AON and 70% standard methods shows the odds of mortality in one group (complete care and at least 70% of the necessary care, respectively) compared to the other group (not complete care and less than 70% of the necessary care, respectively), for example, patients who received at least 70% of the necessary care components have 40.4% lower odds of mortality compared to patients who received less than 70% of the eligible care components, for unadjusted analysis.

Predictive accuracy of composite measures

All types of composite measures performed poorly in predicting 30-day all-cause mortality, assessed with apparent validation and 5-fold cross-validation. (Figure 12, Figure 13) Similarly, composite measures provided very little incremental predictive value over included covariates. ROC curves for all models (models with only covariates, models with only composite measures and models with composite measures and covariates) for apparent validation and 5-fold cross-validation are available in appended paper III.

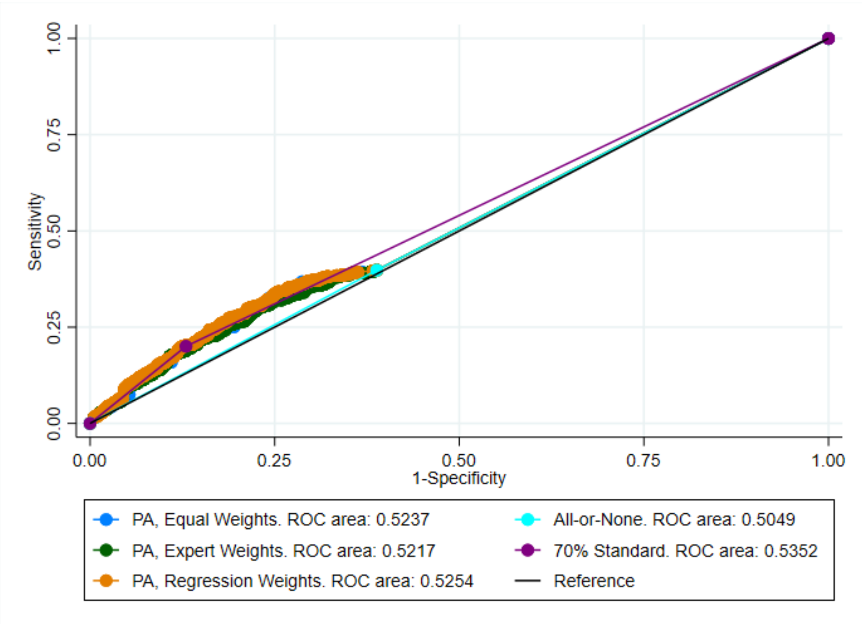

Figure 12. ROC curves for composite measures, apparent validation [53]

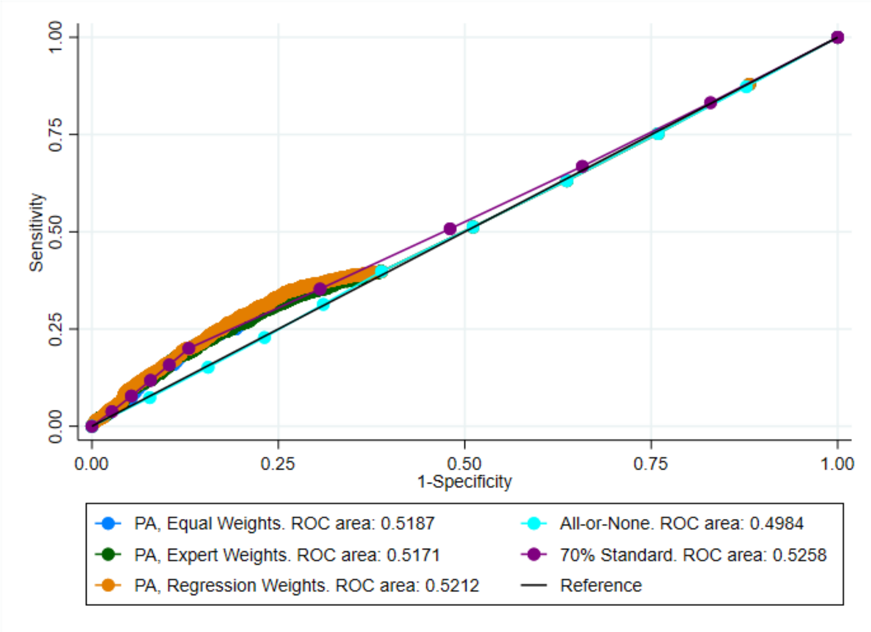

Figure 13. ROC curves for composite measures, 5-fold cross-validation [53]

Sensitivity analysis
Results from the sensitivity analysis were comparable to the results from main analysis. For more detailed information, please see appended paper III. [53]

5.4. DISCUSSION

5.4.1. PAPER II

Results from paper II emphasized the importance of awareness with regards to potential effects of using different approaches to construct composite measures. Using different approaches led to different rankings for the majority of hospitals that cannot be solely explained by random variation. Therefore, when evaluating healthcare providers based on composite score rankings, consideration of such potential uncertainties related to use of composite measures is required. [52]

5.4.2. PAPER III

Results from paper III showed changes in hospital rankings for some of the hospitals, significant association with lower odds of mortality for majority of the methods and poor performance in predicting 30-day mortality for all of the composite measures. These results emphasize the importance of careful interpretation of results obtained by composite measures and importance of awareness regarding to the difference between association and prediction. Poor performance of composite quality scores in predicting mortality should not be considered as an indication of insignificancy of quality of care. Quality of care plays a great role in management of stroke care and patient outcomes, as shown by significant association with lower mortality in this study. However, other factors such as age and stroke severity are stronger predictors of 30-day mortality. Furthermore, the overall high achievement rate for stroke care indicators in Denmark may led to difficulties in discriminating quality of care in healthcare providers. [53]

5.5. METHODOLOGICAL CONSIDERATIONS

The outcomes used in study II and study III were 1-year all-cause mortality and 30-day all-cause mortality, respectively. However, there may be other relevant outcomes that reflects quality of care. As both heart failure and stroke are associated with a great burden for patients, indicators such as quality of life can be especially relevant outcomes of the quality of care.

In study II and study III, selection of included indicators for the composite measures relied on (1) the availability in the DHFR and the DSR, respectively and (2) feedback from the expert panel. There may be other components of care that are important to measure in order to distinguish healthcare providers' performance, but not available in the respective databases or for example, difficult to quantify. Furthermore, different group of experts may select different quality indicators.

Study II and study III were based on observational data and therefore, causation cannot be assumed in neither of these studies. Furthermore, certain types of biases are associated with observational studies.

Selection bias
Selection bias may occur if the procedures that were performed to select subjects and factors affect study participation. [54]

In study II and III, the study populations were identified from national clinical registries. These registries include data from all public hospitals in Denmark and registry of patients in the respective databases are mandatory by law. In Denmark, treatment in public hospitals is free of charge with equal access for all Danish residents. [22] Patients with heart failure and stroke are exclusively admitted to public hospitals. The inclusion and exclusion criteria for registration of patients in the clinical quality registries are the same across regions and hospitals. These ensure all patients with heart failure and stroke should have equal access to healthcare services and healthcare providers must register these patients according to the defined inclusion/exclusion criteria in respective registries. Therefore, the coverage of clinical quality registries is expected to be very high and a systematic exclusion of certain groups of patients assumed to be very limited. We do not expect selection bias to be a concern for study II and study III.

Information bias
Information bias can occur if the collected information about study subjects is incorrect. [54]

In study II and study III, all the used information were obtained from multiple Danish registries. In general, in these registries the data completeness and coverage, and validity of the data are considered to be high. [39, 49, 55-58]

For the clinical quality databases, although guidelines with detailed variable definitions are in place, occurrence of errors during the registration of fulfilment of quality indicators or other variables cannot be completely ruled out. For the DNPR, potential problems such as differences between hospitals regarding coding of diagnoses are reported. [57]

In study II and III, there were considerable amount of missing data for some of the included covariates. We performed multiple imputation by chained equations based on the assumption that the data were missing at random.

Confounding
In study II and study III, we adjusted our analysis according to potential confounders. However, residual confounding cannot be ruled out.

Precision
The statistical precision of study II and study III were supported by relatively large study populations (30,739 and 24,324, respectively) based on Danish registries and the provided 95% CIs. In general, CIs were very narrow for the majority of the composite measures in both of the studies.

Generalizability
It is important to emphasize that in our study the main aim was not to directly express the quality of care provided to HF and stroke patients, but to illustrate use of different types of methodologies to construct composite measures and potential consequences of different approaches, using HF and stroke populations. Therefore, our considerations, methods and overall conclusions should be generalizable and applicable for other countries and healthcare settings.

5.6. CONCLUSION

Results from study II and study III showed that use of different methodologies may have effect on hospital rankings. Although most of the composite measures were associated with lower mortality in both of the studies, study III revealed that none of the composite measures provided a good prediction of 30-day mortality. It is concluded that a careful and thorough construction process and critical interpretation of findings obtained by composite measures are crucial.

CHAPTER 6. OVERALL DISCUSSION

6.1. MAIN RESULTS

The results from this thesis showed that construction of composite quality measures is not a straight-forward process and requires careful evaluation to achieve reliable and meaningful measures of quality. It was seen that using different weighting and aggregation schemes may lead to different hospital rankings and different interpretations of quality in healthcare providers. Therefore, it is important to examine different construction methods and interpret the results obtained from composite measures carefully, with consideration of potential uncertainties related to these measures. Whichever method is preferred, transparency of construction process should be assured, and reader should be informed about potential limitations related to use of composite measures.

6.2. COMPARISON WITH EXISTING LITERATURE

While the number of previous studies that have evaluated the construction process and compared methods was limited, some studies investigated the use of different types of composite measures and its potential consequences.

Ido et al. assessed the quality of stroke care based on The Georgia Coverdell Acute Stroke Registry using AON scoring and weighted patient average approaches. [46] While increase in patient average score was associated with lower odds of 1-year mortality, AON was not associated with mortality. [46]

Eapen et al. examined the effect of AON and overall percentage methods on hospital rankings and correlation with 30-day mortality and 30-day readmission rates for patients with acute myocardial infarction. [59] The results provided similar hospital rankings and correlation with 30-day mortality, whereas neither of the composite measures were correlated with 30-day readmission. [59]

Couralet et al. investigated the change in hospital rankings based on five different types of composite measures of quality of acute myocardial infarction in French hospitals. [15] The results showed that the hospital rankings varied substantially according to the choice of methodology.

Simms et al. constructed three types of composite measures and investigated the impact of composite scoring methodology on hospital rankings and association with 30-day, 6-month and 1-year mortality for patients with acute myocardial infarction. [16] While composite quality indicators were inversely associated with mortality, hospital rankings and magnitude of the association were affected by the choice of methodology. [16]

While the results from the studies included in this thesis were in line with some of the existing literature, for example, substantial change in hospital rankings depending on composite scoring methodology, it would be expected that composite scores' rankability, association with outcome and performance in predicting the outcome will be highly dependent on the population, data structure and included quality indicators. Nevertheless, the main conclusion from the studies in this thesis and existing literature are in agreement: use and interpretation of composite quality measures require great caution and awareness.

CHAPTER 7. PERSPECTIVES

Information needs of the audience

More data and large amount of information may not always translate into more actionable results. On the contrary, it may create information burden for the users/audience. As also suggested in the WHO's quality improvement toolkit, it is important that the presented data is user-friendly. [1] Therefore, one of the important considerations to take into account when communicating the quality of care is the information needs of the audience, whether they need detailed information on each component of quality (presented by individual quality indicators), an "overall picture" of quality (presented by composite measures) or, both of them.

Concerns regarding the loss of information

By the nature of combining multiple indicators that potentially presents different aspect and dimensions of care into a single measure, loss of information can be expected. One of the potential solutions, especially for the quality reports, can be the use of composite indicators not instead of, but as a supplement to individual quality indicators. By providing these two levels of information, user can have a quick understanding of overall quality, but also can have the opportunity to assess the performance on each indicator and where the strengths, weaknesses, and the room for improvement exist in the respective healthcare providers or within the healthcare system.

Underlying data, data structure and quality indicators

The performance of composite measures will be highly dependent on the quality and relevance of underlying data and included indicators. In other words, what goes into composite measures matters. Individual quality indicators can be inspected before inclusion in the composite meases. More critical evaluation and use of the data can ensure whether Donabedian assumptions (improvement in processes of care may lead to improved patient outcomes) hold true for individual quality indicators. Moreover, inclusion of quality indicators should be examined according to the purpose of the composite measure. For example, a quality indicator may be an important measure of quality that is strongly associated with better patient outcomes, but still may be not useful in the composite measure if the main aim is to distinguish between providers and the achievement rates for this indicator is very comparable in all healthcare providers. On the contrary, inclusion of such indicators with very comparable success rates may mask differences in healthcare providers' performances even further.

The most appropriate approach

While there is not a simple and straightforward answer to "which approach provides the best results", the aim and targeted audience can be the important drivers of the construction process.

Different types of composite measures may serve to different purposes. For example, if an outcome can be considered as a "gold-standard", use of regression weights where each quality indicator receives weights proportional to their magnitude of association with this gold-standard end point can be considered. If there is high correlation between individual quality indicators, the use of PCA/FA can be examined. [5] Nevertheless, while these approaches may provide statistically satisfactory results, these results may not translate into the best decisions and policies if the main aim is not to obtain the best model, but for example, to motivate healthcare providers to adhere to certain healthcare policies or guidelines. If we aim to motivate already high-performing healthcare providers for excellence, AON scoring may be a useful scoring methodology. [60] If we aim to investigate improvements in a healthcare provider or system over time, use of methods which reward partial care can bring more insightful and motivating results compared to AON scoring, as AON scoring rewards only complete care and smaller scale improvements may be ignored.

Inclusion of experts such as clinicians or other stakeholders in the construction process of composite measures can be very valuable to achieve meaningful composite measures. However, in some settings, identifying ideal candidates for expert panel, planning and conducting the panel process and reaching a consensus may not be obtainable or can be difficult and time consuming.

The use and presentation of multiple composite scores to the reader can be considered. For example, patient average score and AON scores have very different interpretations of quality. While patient average score shows the percentage of care provided to a patient on average, AON score shows the percentage of patients who received complete care. Presentation of both of these scores can provide a more diverse picture of the quality.

Transparency, the non-negotiable principle

Independent from which construction methods are preferred, transparency should be the non-negotiable principle of construction of composite measures to assure the reader about reliability and credibility of the findings. The reader/audience should be informed about limitations associated with using composite measures and potential uncertainty in rankings obtained by composite measures.

Future steps

Future steps towards better composite measures of quality can include (1) development of a standardized terminology and reporting structure for construction of composite measures which can be useful and facilitate a more transparent communication with the reader and (2) expanding existing guidelines with a special focus on quality of healthcare services and different actors in the healthcare system. Guidelines and toolkits can be utilized as a checklist whenever creating composite measures and can guide various actors in a healthcare system (for example, researchers, healthcare providers and managers) according to their needs and aims (for example, evaluation of improvements over time, comparison on provider or national level). The effect of such initiatives should obviously be examined. Qualitative studies and experiments should be conducted to assess the interpretation and use of composite measures for different stakeholders. The effect of different types of composite measures on for example, behavioural responses of stakeholders and decision-making processes based on the adopted composite scoring approach needs to be investigated.

CHAPTER 8. OVERALL CONCLUSIONS

This thesis included three studies and covered different aspects related to composite measures: (1) use of composite measures in the literature with a special focus to methodological considerations, (2) rankability of composite measures, (3) association between different types of composite measures and clinical outcome and (4) predictive accuracy of composite measures in relation to clinical outcome.

The first study investigated use of composite measures in the literature to assess the process quality of care. Results from study I showed that, although use of composite measures gained attention in the recent years to evaluate the overall quality, attention to methodological aspects of composite measures was limited. Although several studies suggested different construction processes may lead to different results, justification of selected methodology used in reviewed papers were limited and potential limitations of composite measures were not communicated with the reader. It was concluded that while these measures were used increasingly, in depth evaluation of composite measures and their limitations were not common.

Study II and study III investigated the impact of composite scoring methodologies on hospital rankings and size of association with clinical outcome. Results showed that for some of the hospitals, rankings were changed substantially based on the methodology that cannot be solely explained by chance. Improvement in the majority of the composite scores were significantly associated with lower mortality in both of the studies.

Additionally, study III investigated predictive accuracy of composite measures. It was seen that all types of composite measures performed poorly in predicting 30-day mortality.

Overall, in this thesis it was concluded that while composite measures can be very useful and powerful tools of quality measurement, their use and interpretation require a great caution as results obtained by these measures, for example hospital rankings, can be vulnerable to methodological choices. Therefore, to obtain meaningful measures of quality, in-depth considerations regarding certain aspects of composite measures are warranted. Furthermore, clear and transparent communication regards to construction process and methodological choices, and critical interpretation of composite measures are prerequisite to ensure the audience/reader about the accountability of the results.

LITERATURE LIST

[1] World Health Organization. Introducing the WHO quality toolkit: supplemental overview 2022.

[2] World Health Organization Regional Office, for Europe, European Observatory on Health Systems, and Policies, Busse R, Klazinga N, Panteli D, Quentin W. *Improving healthcare quality in Europe: characteristics, effectiveness and implementation of different strategies*. Copenhagen: World Health Organization. Regional Office for Europe, 2019.

[3] Shwartz M, Restuccia JD, Rosen AK. Composite Measures of Health Care Provider Performance: A Description of Approaches. *Milbank Q* 2015;93:788-825.

[4] Peterson ED, DeLong ER, Masoudi FA, O'Brien SM, Peterson PN, Rumsfeld JS *et al*. ACCF/AHA 2010 Position Statement on Composite Measures for Healthcare Performance Assessment: a report of American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures (Writing Committee to Develop a Position Statement on Composite Measures). *J Am Coll Cardiol* 2010;55:1755-66.

[5] OECD, European Commission. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. https://www.oecd.org/sdd/42495745.pdf (28 June 2022, date last accessed).

[6] Reeves D, Campbell SM, Adams J, Shekelle PG, Kontopantelis E, Roland MO. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. *Med Care* 2007;45:489-96.

[7] Donabedian A. Quality Assessment and Assurance: Unity of Purpose, Diversity of Means. *Inquiry* 1988;25:173-92.

[8] Institute of Medicine (US) Committee. *Medicare: A Strategy for Quality Assurance: Volume 1*. Washington (DC): Washington (DC): National Academies Press (US), 1990.

[9] World Health Organization. Handbook for national quality policy and strategy: a practical approach for developing policy and strategy to improve quality of care 2018.

[10] OECD, World Health Organization, World Bank Group. Delivering Quality Health Services: A Global Imperative 2018.

[11] Mainz J. Defining and classifying clinical indicators for quality improvement. *Int J Qual Health Care* 2003;15:523-30.

[12] Donabedian A. Evaluating the quality of medical care. 1966. *Milbank Q* 2005;83:691-729.

[13] Austin PC, Ceyisakar IE, Steyerberg EW, Lingsma HF, Marang-van de Mheen P,J. Ranking hospital performance based on individual indicators: can we increase reliability by creating composite indicators?. *BMC Med Res Methodol* 2019;19:131.

[14] Jacobs R, Goddard M, Smith PC. How robust are hospital ranks based on composite performance measures?. *Med Care* 2005;43:1177-84.

[15] Couralet M, Guérin S, Le Vaillant M, Loirat P, Minvielle E. Constructing a composite quality score for the care of acute myocardial infarction patients at discharge: impact on hospital ranking. *Med Care* 2011;49:569-76.

[16] Simms AD, Batin PD, Weston CF, Fox KAA, Timmis A, Long WR *et al*. An evaluation of composite indicators of hospital acute myocardial infarction care: a study of 136,392 patients from the Myocardial Ischaemia National Audit Project. *Int J Cardiol* 2013;170:81-7.

[17] Barclay M, Dixon-Woods M, Lyratzopoulos G. The problem with composite indicators. *BMJ Qual Saf* 2019;28:338-44.

[18] Kolfschoten NE, Gooiker GA, Bastiaannet E, van Leersum NJ, van de Velde, C J H., Eddes EH *et al*. Combining process indicators to evaluate quality of care for surgical patients with colorectal cancer: are scores consistent with short-term outcome?. *BMJ Qual Saf* 2012;21:481-9.

[19] National Quality Forum. *Composite Performance Measure Evaluation Guidance*. http://www.qualityforum.org (April 2020, date last accessed).

[20] Heale R, Twycross A. Validity and reliability in quantitative studies. *Evid Based Nurs* 2015;18:66-7.

[21] Ministry of Health. *Healthcare in Denmark - An overview*. https://sum.dk/Media/C/A/Healthcare-in%20denmark%20an%20overview%20english-V16-dec.pdf (August 2022, date last accessed).

[22] Mainz J, Kristensen S, Bartels P. Quality improvement and accountability in the Danish health care system. *Int J Qual Health Care* 2015;27:523-7.

[23] Olejaz M, Juul Nielsen A, Rudkjøbing A, Okkels Birk H, Krasnik A, Hernández-Quevedo C. Denmark health system review. *Health Syst Transit* 2012;14:i-192.

[24] OECD (2013), OECD Reviews of Health Care Quality: Denmark 2013: Raising Standards, OECD Reviews of Health Care Quality, OECD Publishing, Paris.

[25] Mainz J, Hansen AM, Palshof T, Bartels PD. National quality measurement using clinical indicators: the Danish National Indicator Project. *J Surg Oncol* 2009;99:500-4.

[26] OECD (2010), Improving Value in Health Care: Measuring Quality, OECD Health Policy Studies, OECD Publishing, Paris.

[27] Uggerby C, Kristensen S, Mackenhauer J, Knudsen SV, Bartels P, Johnsen SP *et al*. From accreditation to quality improvement—The Danish National Quality Programme. *Int J Qual Health Care* 2021;33:mzab071.

[28] Danish Clinical Registries (in Danish, Regionernes Kliniske Kvalitetsudviklingsprogram, RKKP). *Danish clinical quality databases.* . https://www.rkkp.dk (August 2022, date last accessed).

[29] Mainz J, Hess MH, Johnsen SP. The Danish unique personal identifier and the Danish Civil Registration System as a tool for research and quality improvement. *Int J Qual Health Care* 2019;31:717-20.

[30] Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS *et al*. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J* 2016;37:2129-200.

[31] Urbich M, Globe G, Pantiri K, Heisen M, Bennison C, Wirtz HS *et al*. A Systematic Review of Medical Costs Associated with Heart Failure in the USA (2014-2020). *Pharmacoeconomics* 2020;38:1219-36.

[32] Cook C, Cole G, Asaria P, Jabbour R, Francis DP. The annual global economic burden of heart failure. *Int J Cardiol* 2014;171:368-76.

[33] Bragazzi NL, Zhong W, Shu J, Abu Much A, Lotan D, Grupper A *et al*. Burden of heart failure and underlying causes in 195 countries and territories from 1990 to 2017. *Eur J Prev Cardiol* 2021;28:1682-90.

[34] Fitzgerald AA, Allen LA, Masoudi FA. The evolving landscape of quality measurement for heart failure. *Ann N Y Acad Sci* 2012;1254:131-9.

[35] Smith SC,Jr, Fonarow GC, Zhao D. Measuring and Improving the Quality of Heart Failure Care Globally. *JAMA Netw Open* 2020;3:e1918642.

[36] Poon S, Rojas-Fernandez C, Virani S, Honos G, McKelvie R. The Canadian Heart Failure (CAN-HF) Registry: A Canadian Multicentre, Retrospective Study of Inpatients With Heart Failure. *CJC Open* 2022;4:636-43.

[37] Savarese G, Vasko P, Jonsson Å, Edner M, Dahlström U, Lund LH. The Swedish Heart Failure Registry: a living, ongoing quality assurance and research in heart failure. *Ups J Med Sci* 2019;124:65-9.

[38] Heidenreich PA, Hernandez AF, Yancy CW, Liang L, Peterson ED, Fonarow GC. Get With The Guidelines program participation, process of care, and outcome for Medicare patients hospitalized with heart failure. *Circulation: Cardiovascular Quality and Outcomes* 2012;5:37-43.

[39] Schjødt I, Nakano A, Egstrup K, Cerqueira C. The Danish Heart Failure Registry. *Clin Epidemiol* 2016;8:497-502.

[40] Nakano A, Johnsen SP, Frederiksen BL, Svendsen ML, Agger C, Schjødt I *et al*. Trends in quality of care among patients with incident heart failure in Denmark 2003-2010: a nationwide cohort study. *BMC Health Serv Res* 2013;13:391-.

[41] Aho K, Harmsen P, Hatano S, Marquardsen J, Smirnov VE, Strasser T. Cerebrovascular disease in the community: results of a WHO collaborative study. *Bull World Health Organ* 1980;58:113-30.

[42] GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol* 2021;20:795-820.

[43] Feigin VL, Norrving B, Mensah GA. Global Burden of Stroke. *Circ Res* 2017;120:439-48.

[44] Ingeman A, Pedersen L, Hundborg HH, Petersen P, Zielke S, Mainz J *et al*. Quality of care and mortality among patients with stroke: a nationwide follow-up study. *Med Care* 2008;46:63-9.

[45] Bray BD, Ayis S, Campbell J, Hoffman A, Roughton M, Tyrrell PJ *et al*. Associations between the organisation of stroke services, process of care, and mortality in England: prospective cohort study. *BMJ* 2013;346:f2827.

[46] Ido MS, Frankel MR, Okosun IS, Rothenberg RB. Quality of Care and Its Impact on One-Year Mortality: The Georgia Coverdell Acute Stroke Registry. *Am J Med Qual* 2018;33:86-92.

[47] Haas K, Rücker V, Hermanek P, Misselwitz B, Berger K, Seidel G *et al*. Association Between Adherence to Quality Indicators and 7-Day In-Hospital Mortality After Acute Ischemic Stroke. *Stroke* 2020;51:3664-72.

[48] Cadilhac DA, Kim J, Lannin NA, Kapral MK, Schwamm LH, Dennis MS *et al*. National stroke registries for monitoring and improving the quality of hospital care: A systematic review. *Int J Stroke* 2016;11:28-40.

[49] Johnsen SP, Ingeman A, Hundborg HH, Schaarup SZ, Gyllenborg J. The Danish Stroke Registry. *Clin Epidemiol* 2016;8:697-702.

[50] Wildenschild C, Mehnert F, Thomsen RW, Iversen HK, Vestergaard K, Ingeman A *et al*. Registration of acute stroke: validity in the Danish Stroke Registry and the Danish National Registry of Patients. *Clin Epidemiol* 2013;6:27-36.

[51] Kara P, Valentin JB, Mainz J, Johnsen SP. Composite measures of quality of health care: Evidence mapping of methodology and reporting. *PLoS One* 2022;17:e0268320.

[52] Kara P, Valentin JB, Løgstrup BB, Schjødt I, Egstrup K, Vinter N *et al*. Composite measures of quality of care for patients with heart failure in Denmark: A nationwide cohort study [in preparation].

[53] Kara P, Valentin JB, Damgaard D, Vögele P, Rasmussen TS, Murante AM *et al*. Composite measures of quality of care for patients with incident ischemic stroke in Denmark: A nationwide cohort study [in preparation].

[54] Rothman KJ. *Epidemiology*. Oxford: Oxford University Press, 2012.

[55] Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. *Int J Epidemiol* 2017;46:798-798f.

[56] Schmidt M, Schmidt SAJ, Adelborg K, Sundbøll J, Laugesen K, Ehrenstein V *et al*. The Danish health care system and epidemiological research: from health care contacts to database records. *Clin Epidemiol* 2019;11:563-91.

[57] Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol* 2015;7:449-90.

[58] Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol* 2014;29:541-9.

[59] Eapen ZJ, Fonarow GC, Dai D, O'Brien SM, Schwamm LH, Cannon CP *et al*. Comparison of composite measure methodologies for rewarding quality of care: an analysis from the American Heart Association's Get With The Guidelines program. *Circ Cardiovasc Qual Outcomes* 2011;4:610-8.

[60] Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. *JAMA* 2006;295:1168-70.

APPENDICES

Appendix 1: Paper I: Composite measures of quality of health care: Evidence mapping of methodology and reporting.

Appendix II: Paper II: Composite measures of quality of care for patients with heart failure in Denmark: A nationwide cohort study.

Appendix III: Paper III: Composite measures of quality of care for patients with ischemic stroke in Denmark: A nationwide cohort study.

ISSN (online): 2246-1302
ISBN (online): 978-87-7573-826-7
AALBORG UNIVERSITY PRESS