Aalborg Universitet

Health-related quality of life in Denmark on a relative scale

mini-catalogue of mean EQ-5D-3L index scores for 17 common chronic conditions

Olesen, Anne Vingaard; Oddershede, Lars; Petersen, Karin Dam

Published in:
Nordic Journal of Health Economics

Creative Commons License
CC BY 3.0

Publication date:
2016

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

[Link to publication from Aalborg University](#)

Health-related quality of life in Denmark on a relative scale: mini-catalogue of mean EQ-5D-3L index scores for 17 common chronic conditions

ANNE VINGAARD OLESEN [1, 2] *
LARS ODDERSHEDE [1]
KARIN DAM PETERSEN [1]

[1] Center for Healthcare Improvements, Department of Business and Management, Faculty of Social Sciences, Aalborg University, Denmark
[2] The Traffic Research Group, Department of Civil Engineering, Faculty of Engineering and Science, Aalborg University, Denmark

Abstract: In health economic evaluations the quality-adjusted life-year (QALY) is one of the preferred outcome measures. Catalogues of median-based decrements in EQ-5D-3L index scores for chronic conditions exist to inform economic evaluations but may not be appropriate for this purpose as mean, rather than median, EQ-5D-3L index scores are of primary interest. Firstly, we aim to estimate mean decrements in EQ-5D-3L index scores through a simple stratified analysis as an alternative to regression modelling. In addition, we aim to estimate the mean decrement in EQ-5D-3L index scores in percent relative to a disease-free reference population. Secondly, we aim to handle both multiple imputation and appropriate estimation of standard errors in the presence of individual sampling weights. Data on EQ-5D-3L from the National Health Profile, Denmark, 2013, were used to estimate the EQ-5D-3L index scores. Calculation of decrements in EQ-5D-3L index scores of chronic conditions was done while controlling for the additional number of chronic conditions beside the one in question, age and sex. Also, a test of homogeneity of decrements across subgroups was conducted. We provide a mini-catalogue of new percentage-scale decrements in EQ-5D-3L index scores. For example, we estimated that angina was associated with an 8.2% reduction in the EQ-5D-3L index score compared to a reference group without angina. If the mean EQ-5D-3L score was 0.848 among corresponding groups without angina; angina patients would have an EQ-5D-3L index score of $(1-0.082) \cdot 0.848 = 0.778$ using the percentage-scale. The estimated percentage reduction in the EQ-5D-3L index score was homogenous regardless of the number of additional chronic conditions, age and sex. We suggest a percentage-scale estimation of EQ-5D-3L index scores for chronic disorders as an alternative to existing median-based methods. Our estimates stem from a simpler model, which, we argue, is easier to use and interpret.

---

* Correspondence to: Anne Vingaard Olesen, The Traffic Research Group, Department of Civil Engineering, Faculty of Engineering and Science, Aalborg University, Denmark. E-mail: avo@civil.aau.dk

1   Introduction

In health economic evaluations the quality-adjusted life-year (QALY) is one of the preferred outcome measures. The QALY incorporates both quantity and quality of life, thereby allowing for comparison across different diseases, including chronic conditions (Brazier *et al.*, 2007). The most frequently used instrument to derive QALYs in Western countries is the EQ-5D-3L. The EQ-5D-3L is a generic health-related quality of life instrument that has been used to describe population health and health outcomes in clinical trials and health economic evaluations. The EQ-5D-3L index scores are available with country-specific preference weights in 24 countries and regions. In 10 countries including Denmark, the values have been derived through the Time Trade-Off (TTO) elicitation method (Szende *et al., 2014*; Wittrup-Jensen *et al.*, 2009). As EQ-5D-3L index scores are not always readily available, catalogues have been developed with utility scores for chronic conditions, thereby thus making it possible to estimate the loss in health-related quality of life associated with a specific disease of interest (Sullivan and Ghushchyan, 2006; Sullivan *et al.*, 2005; Sullivan *et al.*, 2011).

Health economists prefer mean cost and mean QALYs (and hence mean utilities to estimate mean QALYs) when the result of, for example, an intervention or a prevention programme is presented in a cost utility analysis (CUA) as an incremental cost-effectiveness ratio (ICER) because the mean can be interpreted as a per capita measure of cost per QALY (Brazier *et al.*, 2007). As pointed out by Pullenayegum and colleagues (Pullenayegum *et al.*, 2010), the catalogues referred to above are not based on mean scores but on median scores. Sullivan and colleagues used Censored Least Absolute Deviation regression analysis (CLAD) because it appears that the EQ-5D-3L is not normally distributed and exhibits a significant ceiling effect at 1. Given these factors, CLAD might be the most appropriate method of assessing EQ-5D-3L scores (Sullivan *et al.*, 2005). CLAD assumes that the observed utilities are censored at 1, and hence that the true utility can be greater than 1. However, when the EQ-5D-3L is used for calculation of quality weights conceptually bounded by 1 and -1 then the censoring assumption loses its appropriateness (Pullenayegum *et al.*, 2010). Pullenayegum and colleagues found that when the censoring assumption is not appropriate CLAD is biased – and in fact specified as a median regression (Sullivan and Ghushchyan, 2006, p. 417). Based on the work of Pullenayegum *et al.* we suggest that alternatives are tested and median regression eventually is replaced. Ordinary least squares (OLS) is an unbiased alternative – at least asymptotically (Pullenayegum *et al.*, 2010). But other methods should be tried and results compared. One approach to test is estimation of decrements in the preference-based utility index scores on a relative scale instead of the absolute scale on which decrements typically are calculated. A relative scale corresponds to the ratio of means in contrast to absolute-scale mean differences, that is, the mean EQ-5D-3L score of the selected population with the disease minus the mean EQ-5D-3L score of the corresponding disease-free population.

The aim of the present paper is to estimate a mini-catalogue of percentage-scale mean EQ-5D-3L index scores for 17 common chronic conditions.

2   Methods

2.1   Questionnaire data

For estimation of scores, we utilized data from a health survey answered by 20,220 adults in the North Denmark region in 2013, which included self-reported information on 17 chronic disorders selected independently by the regional health authority as the most important chronic conditions of concern. Initially, the survey was sent to a random sample

of 35,700 adult citizens of the North Denmark region of which 20,200 (56.6 %) were returned. The health survey also included the EQ-5D-3L (Wittrup-Jensen *et al.*, 2009). The EQ-5D-3L comprises five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), each of which has three levels (no problems, moderate problems and extreme problems). An individual's EQ-5D-3L health state can be expressed as a five-digit health profile by combining the levels in each of the five dimensions; this allows a possible 243 health states to be defined (Brooks *et al.*, 2003; Rabin and de Charro, 2001). A single index score can be derived for each of these health states by applying preference weights obtained from the general population (Szende *et al.*, 2007). Danish EQ-5D-3L preference weights have been generated using the TTO valuation technique in a random sample from the Danish general population for some of the 243 possible health conditions and with different regression methods for the remaining conditions (Wittrup-Jensen *et al.*, 2009; Wisløff *et al.*, 2014).

A total of 5,007 of the respondents (25%) had one or more missing answers including missing answers to the five questions of the EQ-5D-3L questionnaire (629 respondents). Therefore, multiple imputation of missing values was performed with the STATA 13 procedure "mi impute chained" (StataCorp, 2013). We chose an imputation model comprising age, sex, educational level and two questions from the questionnaire answered by 99.2% of the study population ("How do you assess your general health?" and information on whether any social benefit was received), thereby assuming that the probability of missing data is conditionally independent of any unobserved factors given knowledge of sex, age, education and answers to the questions mentioned above (Little and Rubin, 2002). Handling of individual non-response weights was done as described in Kreuter and Valliant (Kreuter and Valliant, 2007). The North Denmark region is one of five regions in Denmark and comprises 11 smaller communities. Stratified sampling with regards to communities was performed in this study and the way to handle this stratified approach is also described in Kreuter and Valliant (Kreuter and Valliant, 2007).

2.2   Mean decrements

We aimed to investigate the impact of sex, age and the number of additional chronic conditions beside the one in question on mean decrements. It is generally accepted that the loss of utility associated with chronic conditions is greater among women, the elderly and the chronically ill (Sørensen *et al.*, 2009). We calculated the mean decrements as minus the difference between the mean EQ-5D-3L score among respondents with the chronic condition in question and the mean EQ-5D-3L score among the remaining respondents without the specific condition. Both means were weighted and multiply imputed as described in the previous section and the calculation took place in STATA 13 with the "mean" procedure followed by the "lincom" procedure, thus estimating the standard error of the mean decrement. In order to investigate the effect of sex and age on the mean decrements we stratified or subdivided the data into four strata or disjoint parts (for age: <65 years and 65+ years). Then for each stratum, we estimated a mean decrement and plotted the four numbers per condition. Afterwards we did a similar inspection of the marginal effect of the number of additional conditions beside the one in question (0, 1, 2, 3, 4, 5+ additional conditions); that is, we subdivided the data into six parts, estimated a mean decrement in each stratum and plotted the six numbers per condition for comparison.

2.3   Mean decrements in per cent

Similarly, we estimated stratum-specific decrements on the relative scale on which we compare the ratio of means of the respondents without a disease (nominator) and those with

the disease respectively (denominator). The relative-scale decrement is defined as the ratio of means minus one multiplied by minus one hundred. However, we did the analysis of the ratio on the logarithmic scale. Each stratum-specific ratio of means was transformed with the natural logarithmic function because the asymptotic normality of ratio estimates use to fit better on the logarithmic scale. The estimation of the log ratios in STATA 13 was performed using the "mean" procedure followed by the "nlcom" procedure which utilizes the so-called delta method for calculation of standard errors (Kirkwood and Sterne, 2003, p. 157).

2.4   An alternative to regression modelling: stratified analysis

In epidemiology, stratification by important confounders accompanied by inverse variance weighting is a well-known tool of confounder adjustment (Rothman and Greenland, 1998). The procedure of weighting together different (stratum-specific) estimates is also used in fixed-effects meta-analysis (Kirkwood and Sterne, 2003). In this work, we considered three possible confounders: age, sex, and the number of co-morbid chronic conditions beside the one in question. Because the percentage-scale decrements cannot be estimated in linear regression as the absolute-scale decrements we used the stratified approach here. We now describe the alternative approach in details which is also exemplified in the appendix.

With the aim of adjusting for sex, age (in two categories) and NACC (the number of additional conditions beside the one in question, in six categories), we subdivided the data into 24 strata corresponding to the triple interaction of the three variables. Then we estimated both absolute- and percentage-scale decrements with standard errors in each stratum. Firstly, we considered the adjusted relative-scale decrement. As seen in the appendix, the weights are the inverse squared standard error ($weight = \frac{1}{SE^2}$). The stratified approach rests upon an assumption of homogeneity of effect in all strata, that is, a similar sign and magnitude of decrements in all disjoint subsets of data. This hypothesis can be tested by a Chi-square test as found in Kirkwood and Sterne (Kirkwood and Sterne, 2003):

$\chi^2 = Q = \sum_i weight_i \cdot \ln(ratio_i)^2 - \frac{[\sum_i weight_i \cdot \ln(ratio_i)]^2}{\sum_i weight_i}$, where the index i numbers the strata and the degrees of freedom equal the number of strata minus one.

Secondly, we estimated the common percentage-scale decrement adjusted for sex, age and NACC: the common $\ln(ratio) = \frac{\sum_i weight_i \cdot \ln(ratio_i)}{\sum_i weight_i}$ whereas the standard error was estimated as $SE(\ln(ratio)) = \frac{1}{\sqrt{\sum_i weight_i}}$. Using the exponential function leads us back to the original scale: the common $ratio = \exp(\ln(ratio))$. The exponential function can also be used on the 95% confidence interval of the common $ln(ratio)$ to get a 95% confidence interval of the common $ratio$.

The adjusted estimate of the mean decrement on the absolute scale was conducted similarly through replacement of $\ln(ratio)$ with the mean difference in each stratum. The standard error is now the standard error of the mean difference, i.e. the absolute-scale decrement obtained from STATA 13. The common absolute decrement adjusted for sex, age and NACC $= \frac{\sum_i weight_i \cdot difference_i}{\sum_i weight_i}$ and $SE = \frac{1}{\sqrt{\sum_i weight_i}}$ as above where again the $weight_i = \frac{1}{SE_i^2}$ and SE_i is the standard error of $difference_i$ in the ith stratum. Also the formula of the homogeneity test is straightforward: $\chi^2 = Q = \sum_i weight_i \cdot difference_i^2 - \frac{[\sum_i weight_i \cdot difference_i]^2}{\sum_i weight_i}$, where the term "difference" denotes the absolute-scale decrement. The degree of freedom is equal to the number of strata minus one.

3   Results

Table 1 shows the descriptive statistics of the weighted and imputed data to the left and descriptive statistics of the complete cases to the right. It can be seen from the table that the imputed data contained more respondents with chronic conditions (13,365 (66.1%) versus 9,760 (64.2%)) and that these respondents have lower mean quality of life (total population score weighted and imputed of 0.844 versus a score of 0.865 for the un-weighted complete cases).

After multiple imputation was performed, we calculated weighted mean EQ-5D-3L scores for respondents with a chronic disease and separately for the remaining respondents without the disease for each of the 17 states, respectively. The mean decrement in EQ-5D-3L score was simply the difference between those means. See Table 2 for an example (angina). Here, the mean EQ-5D-3L was calculated for the group of respondents with self-reported angina (=0.648, also found in Table 1) and the rest without angina, respectively (=0.848). We also calculated the ratio of means as seen in the table and transformed the ratio by the natural logarithmic function. The standard errors of the mean decrement and the log-scale mean ratio were obtained from STATA 13.

In Figure 1, we show the mean decrements stratified by sex and age (in two categories: <65 years and 65+ years) for each of the 17 chronic diseases. There seemed to be no effect of sex because the line segments corresponding to a specific disease tended to be horizontal and parallel with respect to age. However, a very strong effect of age can be seen in the figure. For 14 out of 17 diseases for females, the mean decrement decreased or became numerically smaller in the old age category than in the younger group. For the males, 16 out of 17 decrements decreased with age. In Figure 2, we stratify by the number of additional chronic conditions (NACC) beside the condition in question and show again the decrements by NACC (0 additional conditions, 1, 2, 3, 4, 5+). There is a very slight decreasing pattern in decrements by NACC; 12 out of 17 diseases tend to decrease by NACC. Of the remaining five, only two diseases (asthma and cerebral thrombosis) show a strong increasing trend. Only one curve showed a statistically significant effect modification by NACC, namely migraine, which may be interpreted as a random finding because the corresponding curve is rather flat compared with the others.

In Table 3, we report both the absolute- and relative-scale mean decrements originating from a stratification by sex (2 categories), age (2 categories) and NACC (6 categories) followed by an inversely weighted average of the 24 mean decrements on either the absolute or relative scale. The stratification and weighted average calculation corresponds to a linear regression analysis adjusting for the triple interaction of sex, age and NACC for the absolute-scale decrement. A linear regression estimation of relative-scale decrements is not possible and can be calculated only by the stratification method described here. An example of the calculation behind the weighted average is given in the appendix. In the appendix, the three sums in Q are calculated in adjacent columns and finally, used in the estimation of Q=10.29 for angina. The p-value= 0.99 is the probability of obtaining the value $Q = 10.29$ or more in the chi-square distribution with 23 degrees of freedom. On the basis of this large p-value (much larger than 0.05) we cannot reject the null hypothesis of homogeneity of the 24 relative decrements in the case of angina. For all 17 chronic conditions, we obtained p-values above the significance level of 5% when testing the hypothesis of homogeneity of decrements.

Table 1:   **Descriptive statistics of the study population**

Chronic condition	Weighted and multiply imputed					Complete cases unweighted*						
	N	%	mean age	% women	Unadjusted mean EQ-5D-3L	n	%	NCC 25%	NCC 50%	NCC 75%	median EQ-5D-3L	Unadjusted mean EQ-5D-3L
Asthma	1436	7.1	49	60	0.761	889	5.8	2	3	4	0.824	0.802
Allergy	3761	18.6	47	59	0.821	2665	17.5	1	2	3	0.824	0.839
Diabetes	1193	5.9	65	48	0.740	669	4.4	2	3	4	0.818	0.781
Hypertension	4428	21.9	64	50	0.771	3097	20.4	2	2	4	0.824	0.813
AMI	283	1.4	70	33	0.659	139	0.9	3	4	6	0.723	0.686
Angina	465	2.3	65	46	0.648	247	1.6	3	4	6	0.723	0.692
Cerebral thrombosis	364	1.8	68	44	0.621	177	1.2	2	4	5	0.717	0.658
COPD	1072	5.3	65	47	0.688	574	3.8	2	4	5	0.771	0.735
Osteoarthritis	4570	22.6	63	58	0.719	3214	21.1	2	3	4	0.776	0.756
Rheumatoid arthritis	1294	6.4	61	51	0.651	739	4.9	2	4	5	0.756	0.697
Osteoporosis	829	4.1	71	81	0.672	422	2.8	2	3	5	0.774	0.728
Cancer	586	2.9	66	50	0.728	400	2.6	2	3	4	0.776	0.773
Migraine	3073	15.2	46	70	0.742	2161	14.2	1	2	4	0.818	0.780
Psychiatric disorder	1537	7.6	45	61	0.653	913	6.0	2	3	4	0.756	0.686
Slipped disk	2912	14.4	55	49	0.695	1970	12.9	2	3	4	0.776	0.728
Cataracts	1072	5.3	74	59	0.716	548	3.6	2	4	4	0.776	0.775
Tinnitus	2528	12.5	57	34	0.787	1790	11.8	2	2	4	0.824	0.817
Total population	20220	100	49	50	0.844	15213	100	0	1	2	0.824	0.865
No chronic conditions	6855	33.9	40	45	0.934	5453	35.8	0	0	0	1	0.938
At least one condition	13365	66.1	53	53	0.798	9760	64.2	1	2	3	0.824	0.824
One chronic condition	5197	25.7	46	50	0.879	4232	27.8	1	1	1	0.838	0.886
Two	3498	17.3	53	52	0.820	2702	17.8	2	2	2	0.824	0.832
Three	2083	10.3	58	55	0.752	1445	9.5	3	3	3	0.776	0.769
Four	1213	6.0	62	57	0.693	749	4.9	4	4	4	0.756	0.713
Five+	1375	6.8	66	57	0.597	632	4.2	5	5	6	0.708	0.633

* Quartiles cannot be computed using sampling weights. NCC=number of chronic conditions. Note, that the values 0 and 1 are categorized separately.

Note: Descriptive statistics on the study population both before (to the right) and after (to the left) multiple imputations. Number of respondents with a specific chronic condition, prevalence, mean age and EQ-5D-3L, and percentage of women were weighted by individual sampling weights. Quartiles and medians (and mean EQ-5D-3L for comparison) of the number of chronic conditions (NCC) and EQ-5D-3L score are provided for the complete cases. Data originated from 20,220 respondents of a health survey in the North Denmark region in 2013. Missing data were multiply imputed and sampling weights appropriately handled through Taylor series expansions (Little and Rubin, 2002, p.53; Kreuter and Valliant, 2007, p. 9-12).

Table 2:   Estimated mean absolute-scale and mean relative-scale decrements for angina. An example

|	Multiply imputed and weighted mean (SE)	Mean decrement (SE)	Mean ratio	Mean ratio on log-scale (SE)	Relative-scale decrement
Respondents *with* self-reported angina	0.648 (0.0185)	0.200 (0.0186)	0.764	-0.270 (0.0287)	23.6%
Respondents *without* self-reported angina	0.848 (0.0017)	Reference	Reference	Reference	Reference

Note: Mean EQ-5D-3L scores for respondents with and without self-reported angina respectively. The mean decrement in EQ-5D-3L index score is minus the absolute difference between the means similar to what can be estimated as decrements in OLS. The relative-scale decrement is the relative difference between the means i.e. the mean ratio minus one multiplied by minus 100. The table also reports the mean ratio on the logarithmic scale on which the standard error was calculated. Note that both the absolute- and relative-scale decrements reported here are large because no confounder adjustment was made. Data from 20,220 respondents of a health survey in the North Denmark region, 2013.

Figure 1:   Mean decrement in EQ-5D-3L index score by sex and age for 17 common chronic conditions

Note: Estimated mean decrements in EQ-5D-3L index score marginally by the combination of sex and age (in two categories: <65 years and 65+ years) for 17 chronic conditions. The well-known pattern of smaller decrements in the elderly is recognized from this plot. Only a few chronic conditions (2–3 out of 17) show an increasing trend. There are no signs of lower mean decrements for females in this data set after stratification by age. Data from 20,220 respondents of a health survey in the North Denmark region, 2013.

Figure 2:   **Mean decrement in EQ-5D-3L index score by number of additional chronic conditions for 17 common chronic conditions**
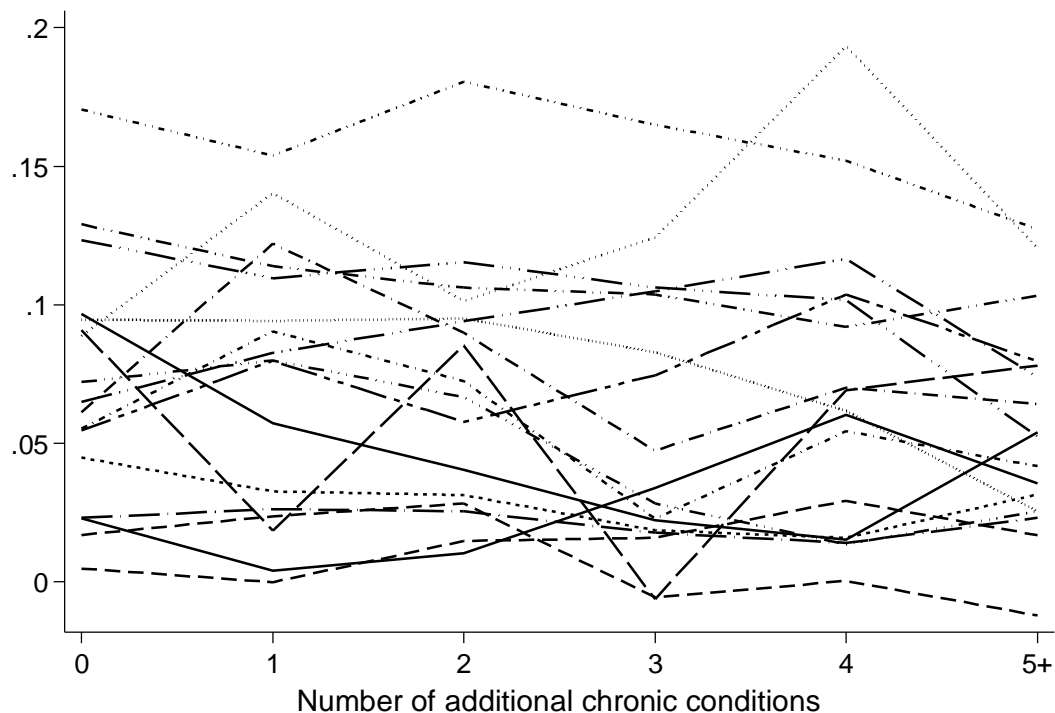
Note: Estimated mean decrements in EQ-5D-3L index score marginally by the number of additional chronic conditions (NACC) beside the condition in question for 17 chronic conditions. The picture is rather flat curves with a slight decreasing pattern by NACC. Twelve curves show a decreasing trend whereas five curves tend to increase. Data from 20,220 respondents of a health survey in the North Denmark region, 2013.

3.1   An example of the application of percentage-scale decrements

We will give an example of the difference between the absolute- and relative-scale decrements and the properties of these two types of application. Let us consider angina in Table 3, for which we find a mean decrement of 0.068 whereas the relative decrement was estimated to be 8.2%. We calculate the mean EQ-5D-3L index score stratified by NACC in the population with no angina and get: 0.934 (NACC=0), 0.879 (NACC=1), 0.820 (NACC=2), 0.753 (NACC=3), 0.692 (NACC=4) and 0.603 (NACC=5+). The absolute mean decrement is interpreted as being a constant 0.068 in all six strata, whereas on the relative scale, the decrement can be calculated as 0.934 multiplied by 0.082, which equals 0.077 (NACC=0), and similarly: 0.072 (NACC=1), 0.067 (NACC=2), 0.062 (NACC=3), 0.057 (NACC=4) and 0.049 (NACC=5+) decreasingly by the number of co-morbid chronic states reflecting the pattern seen in Figure 2. The relative decrements can, in a similar manner, reflect the decreasing pattern seen for age in Figure 1.

The ranking, in Table 3, of the conditions by the absolute magnitude of the decrements seems consistent with the severity of the conditions. The worst states are mental disorder and cerebral thrombosis, whereas allergy and asthma are the least severe conditions according to the percentage-scale. The ranking according to the conventional absolute scale was similar.

Table 3:   Estimated decrements in mean EQ-5D-3L index score on both absolute and relative scale adjusted for sex, age and the number of co-morbid chronic conditions

| | Adjusted by stratification* | | | | | | | |
| | Absolute scale | | | Relative scale | | | | |
Chronic state	mean decrement	95% CI		mean ratio	95% CI		mean decrement in %	95% CI	
Asthma	0.011	-0.002	0.025	0.989	0.973	1.005	1.1%	-0.5%	2.7%
Allergy	0.000	-0.007	0.006	1.000	0.992	1.007	0.0%	-0.7%	0.8%
Diabetes	0.035	0.022	0.048	0.959	0.944	0.974	4.1%	2.6%	5.6%
Hypertension	0.022	0.015	0.030	0.975	0.967	0.983	2.5%	1.7%	3.3%
AMI	0.047	0.019	0.076	0.946	0.911	0.984	5.4%	1.6%	8.9%
Angina pectoris	0.068	0.043	0.092	0.918	0.888	0.949	8.2%	5.1%	11.2%
Cerebral thrombosis	0.114	0.088	0.139	0.873	0.843	0.905	12.7%	9.5%	15.7%
COPD	0.059	0.043	0.076	0.928	0.908	0.948	7.2%	5.2%	9.2%
Osteoarthritis	0.090	0.082	0.098	0.897	0.889	0.906	10.3%	9.4%	11.1%
Rheumatoid arthritis	0.100	0.086	0.115	0.881	0.864	0.899	11.9%	10.1%	13.6%
Osteoporosis	0.071	0.055	0.088	0.918	0.898	0.938	8.2%	6.2%	10.2%
Cancer	0.061	0.044	0.079	0.928	0.908	0.949	7.2%	5.1%	9.2%
Migraine	0.072	0.063	0.081	0.920	0.910	0.930	8.0%	7.0%	9.0%
Mental disorder	0.157	0.143	0.171	0.821	0.806	0.837	17.9%	16.3%	19.4%
Slipped disk	0.106	0.097	0.116	0.875	0.864	0.887	12.5%	11.3%	13.6%
Cataracts	0.048	0.030	0.066	0.947	0.926	0.968	5.3%	3.2%	7.4%
Tinnitus	0.020	0.012	0.028	0.977	0.968	0.986	2.3%	1.4%	3.2%

Note: Estimated decrements in mean EQ-5D-3L index scores on both absolute and relative scale adjusted for sex, age and the number of additional chronic conditions beside the state in question. The calculation was performed using a stratified approach and involved multiple imputation and appropriate handling of individual sampling weights. Data from 20,220 respondents of a health survey in the North Denmark region, 2013. *Stratification was done with respect to the number of additional chronic conditions beside the one in question, sex, and age (in two categories: <65 years and 65+ years) except for AMI, cerebral thrombosis and osteoporosis, which were only adjusted for NACC and age because of scarcity of data in some strata.

4   Discussion

As something new, this paper presents percentage-scale mean decrements of index scores of chronic conditions. These relative decrements may be more homogeneous according to age, sex and the degree of co-morbidity (measured as the number of chronic conditions beside the one in question). Furthermore, the suggested measures are based on mean differences and mean ratios – not median differences as suggested by others (Sullivan *et al.*, 2005; Sullivan *et al.*, 2011). The work of Pullenayegum and colleagues is essential to daily life economic evaluations (Pullenayegum *et al.*, 2010); the solutions offered by this group are, however, cumbersome to implement. Pullenayegum et al. point to latent class models, two-part models and generalized additive models (Pullenayegum *et al.*, 2010; Pullenayegum

et al., 2013) with the aim of modelling the underlying EQ-5D-3L distribution. What we do is simpler but rests heavily on an assumption of the appropriateness of the mean as a measure of the central tendency in the sampling distribution of EQ-5D-3L – an assumption shared with other solutions as well but also an assumption that means less to the economist who wants the mean. The mean and median coincide systematically in symmetric distributions; in skewed multimodal distributions with a considerable probability mass in 1 (like the typical EQ-5D-3L distribution), the mean and median only coincide by chance. Facing this fact implies discontinuation of the use of censored least absolute deviations (CLAD) median regression as argued by Pullenayegum and colleagues (Pullenayegum *et al.*, 2010).

It is recommended that our stratified analysis is repeated in other data sets in order to gather information on the validity and reliability of the method. The suggested stratified analysis facilitates estimation of relative decrements on the percentage scale which could make the work with, for example, Markov models easier when implicit control of the number of additional chronic conditions, age and sex is carried out. The application of decrements from the original catalogues easily becomes ambiguous because one needs to know more about the chronicity of the persons of interest in an evaluation – information that of course in some studies is available. The original median-scale decrements vary by the total number of chronic conditions of the person in question. The mean decrements that we present are invariant, and we have presented a statistical test of the significance of the interaction by the number of additional chronic conditions, age and sex. Testing gave for our 17 chronic conditions only insignificant p-values after stratification by the three variables in 24 strata. We tried to adjust the mean decrements for age and sex but the estimates were not homogeneous by test, and we chose to further stratify by the number of additional chronic conditions, which is a strong confounder. We decided to skip more socio-economic variables in our analysis, not only for simplicity, but also because we claim that inclusion of many socio-economic variables may thin out the utility loss due to a specific chronic condition. Serious disease may affect the socio-economy of the person with the disease and we aim to keep, for example, the utility loss caused by a dropout from the work force and lower income in the estimate, say, of cerebral thrombosis.

New catalogues of preference-based index scores for chronic conditions are required, i.e. estimated using alternative methods that are able to handle both multiple imputation and appropriate estimation of standard errors in the presence of individual sampling weights. Our work meets both requests.

Our mini-catalogue of relative mean decrement EQ-5D-3L index scores only comprises 17 chronic conditions, which, of course, is too little. However, the stratified analysis could be repeated in data with information on more conditions. On the other hand, one might ask whether there exists an upper bound of the number of conditions included in one model, and whether it would be more appropriate to estimate the utility loss of a specific condition from subject matter data exclusively including only subject matter co-morbidity, and thus leaving out all other "competing" chronic conditions. Comparison of the catalogued utility estimates by Sullivan and colleagues (Sullivan *et al.*, 2005; Sullivan *et al.*, 2011; Sullivan and Ghushchyan, 2006) and our estimates is not straightforward because Sullivan and colleagues include the number of chronic conditions in the application of their estimates whereas we do not. What we observed in our results were positive signs of all decrement estimates. Furthermore, we noticed what we interpret as consistency with respect to seriousness of the condition in the ranking of the 17 chronic conditions by magnitude of the estimates.

A limitation of our approach is the lack of a thorough sensitivity analysis of the used imputation model. We reran the calculation of scores with a simple imputation model with

no other variables than the endogenous five questions of the EQ-5D-3L and the 17 chronic conditions. We got a maximal difference between estimates of 0.017 for absolute scale, 2.4 percent points for relative scale estimates and standard errors of similar magnitude as well. Complete case regression analysis (including individual sampling weights) was also performed and resulted in maximal differences of 0.025 and 3.5 percent points respectively for 16 conditions excluding AMI. For AMI a difference of 0.088 was found between absolute estimates, whereas for relative estimates a discrepancy of 9.7 percent points was seen. This larger discrepancy was most likely induced by the lack of correction for sex due to scarcity of data as mentioned in the footnote of Table 3. The standard errors were slightly smaller in the complete data case because the imputation of missing data implies larger variation of estimates.

We should emphasize that the application of the percentage-scale mean decrements in EQ-5D-3L index scores becomes easy because they are homogeneous with respect to sex, age and the number of chronic conditions (beside the one in question). And we stress that decrements on a relative scale are able to reflect the decreasing decrements by, for example, age, which is seen empirically.

5   Conclusion

This study provided percentage-scale mean decrements of index scores of chronic conditions, mean decrements that fit well with empirical research. The decrements were estimated with appropriate handling of both multiple imputation of missing data and corrections of standard errors in the presence of individual sampling weights.

References

Brazier, J., Ratcliffe, J., Tsuchiya, A. and Salomon, J. (2007). *Measuring and valuing health benefits for economic evaluation*. Oxford: Oxford University Press

Brooks, R., Rabin, R., and de Charro, F. (2003). *The measurement and valuation of health status using EQ-5D: A European perspective*. Dordrecht: Kluwer Academic Publishers.

Kirkwood, B. R. and Sterne J. A. C. (2003). *Essential medical statistics*. Malden, MA: Blackwell Science Ltd.

Kreuter, F. and Valliant, R. (2007). A Survey on Survey Statistics: What is Done and Can be Done in Stata. *The Stata Journal*, 7(1), 1-21.

Little, R. and Rubin, D. (2002). *Statistical analysis with missing data (2nd ed)*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Pullenayegum, E., Tarride, J. E., Xie, F., Goeree, R., Gerstein, H. C. and O'Reilly, D. (2010). Analysis of Health Utility data when Some Subjects Attain the Upper Bound of 1: Are Tobit and CLAD Models Appropriate? *Value in Health*, 13(4), 487-494.

Pullenayegum, E., Wong, H. and Childs, A. (2013). Generalized Additive Models for the Analysis of EQ-5D Utility data. *Medical Decision Making*, 33, 244-251.

Rabin, R. and de Charro, F. (2001). EQ-5D: A Measure of Health status from the EuroQoL Group. *Annals of Medicine*, 33(5), 337-343.

Rothman, K. and Greenland, S. (1998). Introduction to Stratified Analysis. In: *Modern epidemiology* (pp. 253-279). Philadelphia, PA: Lippincott Williams & Wilkins.

StataCorp (2013). *STATA statistical software: Release 13*. College Station, Texas: StataCorp LP.

Sullivan, P. and Ghushchyan, V. (2006). Preference-based EQ-5D Index Scores for Chronic Conditions in the United States. *Medical Decision Making*, 26, 410-420.

Sullivan, P., Lawrence, W. and Ghushchyan, V. (2005). A National Catalog of Preference-based Scores for Chronic Conditions in the United States. *Medical Care*, 43, 736-749.

Sullivan, P., Slejko, J., Sculpher, M. and Ghushchyan, V. (2011). Catalogue of EQ-5D Scores for the United Kingdom. *Medical Decision Making*, 31, 800-804.

Szende, A., Janssen, B. and Cabases, J. (2014). *Self-reported population health: An international perspective based on EQ-5D*. Dordrecht: Springer International.

Szende, A., Oppe, M. and Devlin, N. (2007). *EQ-5D value sets: Inventory, comparative review and user guide*. Dordrecht: Springer Verlag.

Sørensen, J., Davidsen, M., Gudex, C., Pedersen, K. M. and Brønnum-Hansen, H. (2009). Danish EQ-5D Population Norms. *Scandinavian Journal of Public Health*, 37, 467-474.

Wisløff, T., Hagen, G., Hamidi, V., Movik, E., Klemp, M. and Olsen, J. A.. (2014). Estimating QALY Gains in Applied Studies: A Review of Cost-Utility Analyses Published in 2010. *PharmacoEconomics*, 32(4), 367-375.

Wittrup-Jensen, K., Lauridsen, J., Gudex, C. and Pedersen, K. (2009). Generation of a Danish TTO Value Set for EQ-5D Health States. *Scandinavian Journal of Public Health,* 37(5), 459-466.

Appendix

Table A1:   Calculation of percentage-scale decrements using stratified analysis

Sex	age	NACC	mean EQ-5D-3L + angina	-angina	ratio	ln(ratio)	SE(ln(ratio))	weight	weight*ln(ratio)	weight*ln(ratio)^2
M	<65	0	0.86	0.94	0.92	-0.08	0.07	178.27	-14.99	1.26
M	<65	1	0.72	0.89	0.81	-0.21	0.07	183.88	-38.07	7.88
M	<65	2	0.70	0.83	0.84	-0.17	0.09	126.43	-21.74	3.74
M	<65	3	0.69	0.73	0.94	-0.06	0.16	41.50	-2.59	0.16
M	<65	4	0.60	0.70	0.85	-0.16	0.21	21.85	-3.49	0.56
M	<65	5+	0.46	0.60	0.77	-0.26	0.24	17.65	-4.61	1.21
M	65+	0	0.88	0.94	0.94	-0.06	0.07	206.58	-13.30	0.86
M	65+	1	0.71	0.90	0.79	-0.23	0.13	59.71	-13.85	3.21
M	65+	2	0.79	0.83	0.95	-0.05	0.05	360.81	-18.14	0.91
M	65+	3	0.75	0.80	0.93	-0.07	0.05	363.87	-25.14	1.74
M	65+	4	0.65	0.76	0.86	-0.15	0.10	96.87	-14.30	2.11
M	65+	5+	0.60	0.66	0.92	-0.09	0.07	182.41	-15.95	1.39
F	<65	0	0.91	0.93	0.98	-0.02	0.08	159.14	-3.43	0.07
F	<65	1	0.84	0.87	0.97	-0.04	0.08	175.07	-6.18	0.22
F	<65	2	0.75	0.81	0.92	-0.08	0.05	374.69	-29.26	2.28
F	<65	3	0.65	0.74	0.88	-0.12	0.20	24.74	-3.08	0.38
F	<65	4	0.58	0.63	0.92	-0.09	0.25	16.19	-1.42	0.12
F	<65	5+	0.47	0.56	0.83	-0.19	0.17	33.93	-6.36	1.19
F	65+	0	0.84	0.93	0.91	-0.09	0.07	187.00	-17.31	1.60
F	65+	1	0.83	0.87	0.96	-0.04	0.13	61.12	-2.32	0.09
F	65+	2	0.66	0.80	0.83	-0.19	0.10	92.39	-17.68	3.38
F	65+	3	0.71	0.76	0.94	-0.06	0.08	162.43	-9.23	0.52
F	65+	4	0.65	0.70	0.93	-0.08	0.09	116.83	-8.93	0.68
F	65+	5+	0.59	0.60	0.98	-0.02	0.07	219.32	-5.40	0.13
Sum								3462.68	-296.76	35.72

Q=	10.29
df=	23
p homogeneity=	0.99
Estimated ln(ratio)=	-0.086
SE=	0.017
95% CI= -0.119	-0.052
Estimated ratio=	0.918
95% CI= 0.888	0.949
Estimated rel decrement=	8.21%
95% CI= 5.10%	11.22%

Note: This table exemplifies (for the chosen chronic condition angina) the estimation of percentage-scale mean decrements in EQ-5D-3L index score that is described in more detail in the Methods' section.