Bayesian Hierarchical Models for Meta-Analysis of Quality-of-Life Outcomes

Citation for published version (APA):

Schmitz, S., Makovski, T. T., Adams, R., van den Akker, M., Stranges, S., & Zeegers, M. P. (2020). Bayesian Hierarchical Models for Meta-Analysis of Quality-of-Life Outcomes: An Application in Multimorbidity. *Pharmacoeconomics*, 38(1), 85–95. <https://doi.org/10.1007/s40273-019-00843-z>

Document status and date:

Published: 01/01/2020

DOI:

[10.1007/s40273-019-00843-z](https://doi.org/10.1007/s40273-019-00843-z)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

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

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

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Bayesian Hierarchical Models for Meta-Analysis of Quality-of-Life Outcomes: An Application in Multimorbidity

Susanne Schmitz¹ · Tatjana T. Makovski^{2,3,4} · Roisin Adams⁵ · Marjan van den Akker^{3,6,7} · Saverio Stranges^{2,8,9} · Maurice P. Zeegers⁴

Published online: 4 October 2019
© Springer Nature Switzerland AG 2019

Abstract

Background Health-related quality of life (HRQoL) is a key outcome in cost-utility analyses, which are commonly used to inform healthcare decisions. Different instruments exist to evaluate HRQoL, however while some jurisdictions have a preferred system, no gold standard exists. Standard meta-analysis struggles with the variety of outcome measures, which may result in the exclusion of potentially relevant evidence.

Objective Using a case study in multimorbidity, the objective of this analysis is to illustrate how a Bayesian hierarchical model can be used to combine data across different instruments. The outcome of interest is the slope relating HRQoL to the number of coexisting conditions.

Methods We propose a three-level Bayesian hierarchical model to systematically include a large number of studies evaluating HRQoL using multiple instruments. Random effects assumptions yield instrument-level estimates benefitting from borrowing strength across the evidence base. This is particularly useful where little evidence is available for the outcome of choice for further evaluation.

Results Our analysis estimated a reduction in quality of life of 3.8–4.1% per additional condition depending on HRQoL instrument. Uncertainty was reduced by approximately 80% for the instrument with the least evidence.

Conclusion Bayesian hierarchical models may provide a useful modelling approach to systematically synthesize data from HRQoL studies.

1 Introduction

Health economic evaluations such as cost-utility analysis (CUA) play an important role in informing reimbursement decisions in many jurisdictions [1, 2]. A CUA compares the incremental cost of a programme with its incremental health benefit to determine the overall value of the programme. Health benefit is typically measured in the form of quality-adjusted life-years (QALYs), taking into account quality as well as quantity of life [2]. Many measures exist to assess quality of life (QoL), and health-related quality of life (HRQoL) in particular. One can distinguish between psychophysical measures, for example the 36-item short form

survey (SF-36) [3] or the WHOQoL-BREF, a 26-item QoL instrument developed by the World Health Organisation [4], and preference-based measures, such as the EQ-5D index value developed by EuroQol [5] or the SF-6D [6]. Psychophysical instruments measure QoL as perceived by the individual, while preference-based instruments use population preferences between health states to value an individual's QoL.

In most countries, the preferred utilities used in CUAs are societal preferences elicited from hypothetical scenarios/health states described by HRQoL measures, such as the EQ-5D, SF-6D or the Health Utilities Index (HUI) [7]. Health states are usually defined by the level of performance ability in a number of health dimensions. HRQoL instruments differ in the number of dimensions and levels used to define health states, as well as in the methodology applied to elicit preferences between those states. For example, the EQ-5D considers five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) with three levels each (no problems, some problems, unable

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40273-019-00843-z>) contains supplementary material, which is available to authorized users.

✉ Susanne Schmitz
susanne.schmitz@lih.lu

Extended author information available on the last page of the article

Key Points for Decision Makers

Bayesian hierarchical models provide an alternative approach to synthesising evidence from utility studies. Such models systematically include evidence from studies applying different assessment methodologies, while borrowing strength across instruments may reduce uncertainty on instrument level.

Reduced uncertainty in instrument-level estimates can reduce decision uncertainty in cost-utility analyses supporting reimbursement decisions. This is particularly beneficial where little evidence exists for the instrument of interest.

Multimorbidity affects quality of life; our model estimates a reduction in quality of life in patients suffering from multimorbidity of 3.8–4.1% per additional disease, depending on the quality-of-life instrument.

to perform), resulting in a total of 243 health states (EQ-5D-3L); preferences are elicited using time trade-off [8]. A new version of the EQ-5D has now been developed allowing for five levels (no, slight, moderate, severe or extreme problems) in each dimension, defining 3125 unique health states (EQ-5D-5L) [9]. The SF-6D is based on the generic SF-36 or the abbreviated 12-item SF-12 profile measures [6, 10]. It considers six dimensions (physical functioning, role limitations, social functioning, pain, mental health and vitality), allowing for four to six levels in each dimension, defining a total of 18,000 health states, preferences for which standard gamble is the chosen elicitation method [11].

Selecting relevant data to inform utility parameters for CUA is of key importance as outcomes are often sensitive to these parameters. The choice of instrument may depend on a preferred measure by the decision maker, data availability, or the condition under investigation. Where HRQoL data were collected alongside major clinical trials informing effectiveness in the model, those may provide the most accurate estimates for the population under investigation. Mappings or crosswalks may be applied, where the collected outcome is not the preferred outcome of interest [12].

Regardless of which outcome scale is chosen, utility input parameters should be evidence-based and informed by all relevant evidence. Many jurisdictions therefore request a systematic review of the literature to identify additional data sources, including the National Institute for Health and Care Excellence (NICE) in the UK, and the National Centre for Pharmacoeconomics (NCPE) in Ireland [13, 14]. A guide on performing a systematic search for health state utilities can be found in the review by Ara et al. [15].

Systematic reviews and meta-analyses (MAs) have become standard tools to inform evidence-based healthcare decisions, and are often deemed to be the highest level of evidence [16]. However, MA of utility studies are less common and remain difficult, mainly due to the high heterogeneity between studies caused by inherent differences in the populations studied, as well as the methodology used to derive utility measures. A recent publication by Petrou and colleagues highlights the challenges and limitations of synthesising evidence of health state utilities [17]. They suggest the use of mixed effects meta-regression models to identify factors influencing utility estimates, in addition to standard fixed or random effects MA. Nevertheless, these methods are still limiting the inclusion of studies to those using comparable methods to derive utility values.

With this paper, we wish to illustrate how Bayesian hierarchical models can provide an alternative approach to synthesising evidence on utility parameters, allowing for the systematic inclusion of data from different preference-based utility measures. While acknowledging differences between alternative preference systems, such models utilise correlations between outcomes and thereby allow for borrowing strength across the entire evidence base. Our model will be illustrated using a case study in multimorbidity (MM). MM is usually defined as the co-occurrence of two or more chronic conditions within an individual [18]. The prevalence of MM in the general population increases rapidly with age and estimates vary largely. Most studies estimate the prevalence in a population aged 60 years and older as above 50%, and up to 99% in some cases [19–22]. MM poses a significant burden on patients and society, with poor QoL together with disability and high healthcare costs [23].

A recent comprehensive systematic review identified studies evaluating the effect of MM on QoL [24]. Frequentist MA for individual QoL instruments confirmed the negative impact, estimating a decline ranging from 1.6% in QoL per additional disease for the mental component score (MCS) of the SF questionnaire (pooled across versions SF-36, SF-12, SF-8) and 4.4% for the physical component of the World Health Organisation's QoL instrument (WHOQoL-BREF). Generally, physical dimensions of wellbeing appeared to be more strongly affected by additional diseases than mental dimensions. The analysis presented in this paper is based on the data extracted from the review by Makovski et al. [24].

Section 2 of this paper provides a brief theoretical overview of alternative approaches to MA, with a focus on HRQoL studies. Section 3 describes the methods used for our case study, the results of which are displayed in Sect. 4. A discussion on the conclusions, advantages and limitation is given in Sect. 5.

2 Meta-Analysis of Utility Data

MA has been developed to combine outcomes from several randomised controlled trials investigating the effect of interest to obtain an overall effect estimate [25]. Motivation behind an MA is to not only increase power and improve precision through increased sample size of the combination of individual studies, but also to obtain an overall value measuring the effect incorporating all available evidence. Many articles and books explain the theory behind MA; a nice overview is given, for example, by Deeks et al. [26]. All approaches estimate an overall effect size as a pooled estimate of the study-level effects, either assuming no systematic differences between trials (fixed effects assumption [FE]) or allowing for between-trial heterogeneity (random effects assumption [RE]). A common approach of weighing the individual studies is the inverse-variance method, which can be adjusted to incorporate between-trial variance for an RE model [27]. Standard MA models can be extended to models that are more complex. Network meta-analysis (NMA) allows for the estimation of relative effects in a network of multiple treatments utilising direct and indirect evidence [28, 29]. Meta-regression allows for the adjustment for heterogeneity by trial-level covariates [30]. While useful for the explanatory investigation of heterogeneity, outcomes of meta-regression should be interpreted with care due to its observational nature (even when conducting an MA of randomised controlled trials) and the risk of ecological fallacy. Both, FE and RE models can be extended to meta-regression; the extension of RE models is also known as mixed-effects meta-regression due to the dual explanation of heterogeneity by RE and covariate information [17].

Bayesian MA models are a widely used alternative to frequentist models. Improvements in computational power have led to the development of very flexible models, which are commonly used, especially in the context of more complex modelling such as NMA [29, 31, 32]. The incorporation of multiple sources of evidence fits naturally within the Bayesian framework, which builds on the incorporation of existing or prior knowledge. In contrast to frequentist models, Bayesian models incorporate prior knowledge in the form of a prior distribution, and communicate outcomes in the form of probability statements. Non-informative prior distributions are commonly used to limit the impact of choice of prior distributions. Frequentist and Bayesian FE and RE MA models usually yield very similar results in this case. Nevertheless, an informative prior distribution can be of interest in the context of MA, in situations where few trials limit the power to accurately estimate between-trial variance in an RE model. Bayesian MA models can be easily extended to meta-regression, explaining heterogeneity caused by potential confounders, similar to the frequentist

approach [30]. Three-level hierarchical models have been previously applied to incorporate evidence sources of various designs [33, 34].

The applicability of standard MA methods in synthesising evidence from multiple utility studies is limited to the incorporation of studies that have applied similar assessment methods [17]. Naïve pooling across different measures cannot be recommended since comparative studies have shown the existence of differences between instruments (see, for example [12, 35–37]). In practice therefore, many evaluations are based on a single or very few utility studies. Especially in situations where little data are available for the chosen outcome measure, this may result in high variance of utility input parameters and therefore increased uncertainty in CUA results. Petrou et al. suggested the use of meta-regression to identify methodological factors with significant impact on utility estimates to either inform subgroup analyses or rely on predicted values of the meta regression [17].

We argue that a Bayesian three-level hierarchical model provides a useful alternative to synthesising utility data. Such a model introduces an outcome scale level between the two standard levels of MA models (the trial and overall levels), equivalent to the study type level reported by Schmitz et al., where such models were applied for the incorporation of evidence of different trial designs [34]. Estimates from different outcome scales are combined using an RE assumption. This allows for differences between scales, while assuming that mean outcome scale estimates are drawn from a common normal distribution. The precision of this distribution is driven by the data itself, with increased precision with decreasing differences between outcome scales. This way, outcome-level estimates benefit from borrowing strength across different outcomes yielding increased precision. Estimates are also drawn towards the overall mean QoL across instruments. The level of borrowing strength depends on the between-outcome heterogeneity. The approach can be especially useful where little information is available for the outcome of interest and where lack of individual patient-level data does not allow for a mapping from alternative outcomes.

3 Methods

3.1 Data

A systematic search of the published literature was conducted to identify eligible studies reported according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [38]. The first search was carried out in November 2016 and last updated in September 2018. Studies were considered for MA if they either reported a linear regression relating a validated HRQoL tool

to MM as defined by the number of conditions, or if reported data allowed for the calculation of a linear slope parameter. Studies with index diseases or specific disease patterns were excluded. Further details on the systematic review, including the PRISMA diagram and search strategy, are published elsewhere [24]. For the purpose of the present analysis, inclusion is further restricted to studies reporting preference-based HRQoL measures, or studies reporting the MCS and the physical component score (PCS) summarising outcomes of the SF-36 (or SF-12). Studies reporting PCS and MCS were further excluded if information on mean age and proportion of female participants is not reported, as these are needed for the transformation to a preference-based scale, as described in the next section. Relevant data extracted for the present analysis include title and author, mean age of the study population, proportion female and mean number of diseases per subject. If not reported, we calculated the mean number of diseases based on MM group sizes. In some instances, this resulted in a lower bound estimate since the most severe group of MM was defined as ‘x diseases or more’, while our estimation assumes all subjects to suffer from ‘x’ diseases in these cases. We extracted the mean and standard error of the slope parameter relating HRQoL and MM. Slopes adjusted for confounders were selected where possible, otherwise unadjusted slopes were recorded. Where neither was reported, we extracted the mean HRQoL, associated variance and number of patients for each MM group in order to estimate a slope ourselves.

3.2 Statistical Analysis

3.2.1 Slope Estimation

Where no slope relating HRQoL to the number of diseases was reported, we estimated a slope based on the data available using weighted linear regression, where possible. Using the HRQoL reported in each MM group (as defined by the number of conditions) and the number of patients in each group as weights, we obtained an estimate of the slope. Where the number of patients per group was not reported, we used the inverse variance as weights. Where data were provided for broader groups (e.g. 0–2 conditions, 3–5 conditions, etc.), we used the midpoint. To adjust for the small degrees of freedom in the group regression, we retrieved the standard error of the slope parameter using the within-group variation. Analyses were conducted using R and Microsoft Excel® [39].

3.2.2 SF-36 Transformation

Many studies identified in the literature review used the SF-36 questionnaire, or the shorter SF-12 version, to assess HRQoL [3]. The SF-36 provides a useful descriptive

measure of HRQoL, however it is not based on a population’s preferences and hence cannot be used in CUAs. However, the SF-6D is a common preference-based measure of HRQoL and uses information from the SF-36 for its health state definitions. It was first developed by Brazier et al. for the SF-36 questionnaire, and later followed a version for the shorter SF-12 [6, 10]. Most studies reporting the SF-36 or SF-12 provide two summary scores, the PCS and the MCS. We applied an algorithm developed by Hanmer to predict average SF-6D scores based on these summary scores, to allow for the inclusion of studies reporting SF-36 or SF-12 summaries in the MA [40]. No such algorithm exists for the SF-8, therefore studies reporting the SF-8 were excluded. The algorithm predicts the average SF-6D based on mean age, proportion female, and MCS and PCS scores:

$$\text{SF-6D} = -0.06446 - 0.00328 \times (\text{Female}) + 0.00012 \times (\text{age}) \\ + 0.00946 \times (\text{MCS}) + 0.00781 \times (\text{PCS}),$$

where MCS and PCS values were reported by the MM group (i.e. separate for each number of diseases); we transformed each individual value and fitted a weighted regression as for the other outcomes. Where only the slope for MCS and PCS was reported, we transformed fitted values based on a reference value of 50, and fitted a slope using weighted regression. The reference value of 50 was chosen as MCS and PCS values in the general US population are normed to have a mean of 50; however, as we are only interested in the decline and not in absolute values, any reference value could be chosen. Since within-group variances could not be recovered for transformed values, the standard error for the transformed slope was assumed to be the same as the maximum standard error of the untransformed slopes directly relating MCS or PCS (normalised to 0–1) with disease count.

3.2.3 Meta-Analysis

Bayesian MA models were fitted in WinBUGs using the R2WinBUGs package in R [39, 41, 42].

In a first step, we fitted individual models for each preference-based HRQoL outcome. Predicted SF-6D studies using the Hanmer algorithm were analysed separately from those reporting SF-6D values in the publication directly. All HRQoL measures were normed to the same scale of 0–1. The outcome of interest was the slope relating HRQoL outcome to number of diseases. We assumed a normal likelihood for the slope parameter and fitted a random effects model to allow for the heterogeneity between studies. A normal prior distribution with a mean of 0 and precision of 0.01 ($N(0, 0.01)$) for the slope parameter, and a uniform prior on the interval (0, 2) [$U(0, 2)$] for the between-study standard deviation (SD), were assumed to be non-informative. Results were displayed as posterior distributions and summarised in terms of mean and SD.

Three-level hierarchical models have been used in the past, for example to combine data from different trial designs [34, 43]. Such models explicitly allow for differences between groups of studies using a random effect on these differences. We applied such a model to allow for the analysis across all preference measures within one model. This allows for borrowing strength across the different measures, thereby reducing uncertainty for measures with little evidence. While the model allows for differences between the outcome measures, a certain similarity is assumed as per the random effects assumption. While the separate analyses pose an independent non-informative prior on the respective outcomes, the slopes estimated for each outcome in the three-level model are assumed to all come from a normal distribution, where the variance describes the heterogeneity between the outcomes. We used a non-informative prior for the heterogeneity between outcomes ($N(0, 0.01)$). Therefore, the effect of borrowing strength across the dataset will be weaker where different outcomes provide rather different results, and stronger where results are more similar across outcomes. The complete WinBUGs code is available as electronic supplementary material (ESM) 1. We display the posterior distribution of the slope parameters, both on HRQoL measure level and overall level, and further summarise outcomes using mean and credible intervals.

One could argue that the model underestimates the uncertainty in estimates derived by the Hanmer algorithm. Through variance inflation, it is possible to adjust weights given to particular outcomes. In addition to the unadjusted analysis, we also present results down-weighting the evidence derived from SF-36/12 studies by 30%.

For completeness, we also fitted frequentist MAs using the metafor package in R [44].

4 Results

4.1 Study Details

The literature review identified 74 studies evaluating QoL in patients with MM, 39 of which were included in the frequentist MA reported by Makovski et al. [24]; 35 were excluded from the quantitative analysis due to insufficient outcome data or insufficient number of studies reporting on the same MM and QoL measure. Of these 39 studies, 31 studies were conducted in the non-hospital setting, measured MM as disease count and used either a preference-based QoL measure or reported MCS and PCS summary scores for the SF-36 or SF-12, together with information on mean age and proportion female. These 31 studies were included in the analysis of this article; study details are summarised in Table 1. Note that one study is included twice as it reports on both SF-6D and EQ-5D outcomes [45]. ESM 2 and 3 detail the PRISMA

flow diagram, as well as reasons for exclusion in the current analysis.

Three studies report SF-6D as an outcome directly, 19 studies report EQ-5D, and, for 10 studies, we transformed MCS and PCS data to SF-6D scales using the Hanmer transformation described above. The smallest study analysed 37 subjects, while the largest study included 831,537 subjects. Covariates show high heterogeneity between the studies: average age ranged from 40 to 94 years, the proportion of females ranged from 49 to 100% (except for one study with 4% female participants [46]) and the average number of diseases per subject ranged from 0.4 to 7. This high variation is also observed within each HRQoL scale, indicating a similar heterogeneity for all HRQoL instruments.

4.2 Statistical Analysis

In a first step, all three outcome measures (SF-6D, EQ-5D and SF-36/SF-12) have been analysed separately using standard Bayesian MA models. The results of these individual analyses are summarised in the first row of Table 2 (independent effects) and Fig. 1a.

All QoL scales indicate a decline in QoL as the number of conditions increase. The largest effect is observed for EQ-5D, indicating a decline of 4.17% per additional disease (slope -0.0417 [SD 0.008]). Estimates based on the transformed SF-36/12 summaries and those directly reporting SF-6D are very similar, indicating a decline of 3.62% and 3.57%, respectively (slope SF-36/12 -0.0362 [SD 0.012], and SF-6D -0.0357 [SD 0.160]). This similarity in outcomes somewhat validates the transformation proposed by Hanmer et al. to provide valid estimates of SF-6D based on SF-36/12 summary measures. While EQ-5D and SF-36/12 summaries estimate slopes, which are significantly negative, the SF-6D slope is not significant due to the increased uncertainty caused by the smaller number of studies included in this analysis. This reduced precision becomes obvious when looking at the posterior distributions in Fig. 1a.

In a second step, all studies have been analysed jointly in the previously described three-level hierarchical model. The results are summarised in Table 2 (row 2, shrunken effects), as well as Fig. 1b, c. Point estimates have changed slightly; the EQ-5D estimate now indicates a decline of 4.11% per additional disease, the SF-36/12 transformed data indicates a decline of 3.77%, and the SF-6D indicates a decline of 3.96%. Compared with the independent effects, these results are shrunken towards the overall average; shrinkage is strongest for SF-6D, where the least evidence is provided. Borrowing strength by jointly analysing all studies also results in a reduction in uncertainty for all QoL scales; again, the effect is strongest for SF-6D, where the SD of the slope reduced over 80%, from 0.16 in the independent analysis to 0.029 in the joint analysis. The effect of shrinkage and

Table 1 Study demographics

	Authors	QoL outcome	Slope (SE)	<i>N</i>	Age (mean)	Female (%)	No. of diseases (mean)	Source slope
1	Barra et al. [45]	SF-6D	− 0.0224 (0.0003) ^a	39,817	45.4	54.5	3.3	Pub
2	Hanmer et al. [47]	SF-6D	− 0.0247 (NA) ^b	95,195	72.3	54.4	NA	GR
3	Lawson et al. [48]	SF-6D	− 0.0736 (NA) ^c	7054	72.7	NA	0.7	GR
4	Agborsangaya et al. [49]	EQ-5D	− 0.0438 (0.0016)	4752	47.7	55.7	1.3	GR
5	Agborsangaya et al. [50]	EQ-5D	− 0.0428 (0.0023)	4946	46.6	52.3	0.7	GR
6	Alonso-Morán et al. [51]	EQ-5D	− 0.0257 (0.0028)	1125	NR	82.1	4.7	Pub
7	Barra et al. [45]	EQ-5D	− 0.0271 (0.0004)	39,817	45.4	54.5	3.3	Pub
8	Ferrer et al. [52]	EQ-5D	0.013 (0.024)	37	94.3	68.0	NA	Pub
9	Gerber et al. [53]	EQ-5D	− 0.0230 (0.0067)	104	77.0	72.1	2.6	GR
10	Heyworth et al. [54]	EQ-5D	− 0.079 (0.0051) ^a	4836	47.9	54.9	0.6	Pub
11	Hodek et al. [55]	EQ-5D	− 0.0120 (NA) ^c	2120	76.3	53.7	1.8	Pub
12	Kim et al. [56]	EQ-5D	− 0.0281 (NA) ^c	1419	72.4	60.7	3.9	GR
13	Lang et al. [57]	EQ-5D	− 0.0960 (NA) ^c	3058	53.4	100	1.1	GR
14	Li et al. [58]	EQ-5D	− 0.1095 (0.0010)	27,806	NA	56.3	1.5	GR
15	Mujica-Mota et al. [59]	EQ-5D	− 0.1185 (NA) ^c	831,537	NA	51.0	0.9	GR
16	Quah et al. [60]	EQ-5D	− 0.0185 (0.0011)	498	73.9	49.0	3.6	GR
17	Rabadi and Vincent [46]	EQ-5D	− 0.0132 (0.0017)	170	69.4	4.0	4.8	GR
18	Sakthong et al. [61]	EQ-5D	− 0.0200 (NA) ^b	1156	50.4	52.0	2.3	GR
19	Sullivan et al. [62]	EQ-5D	− 0.0364 (NA) ^c	47,178	45.0	52.0	1.8	GR
20	Vogel et al. [63]	EQ-5D	− 0.05 (0.0102)	103	70.0	53.4	4.2	Pub
21	N'Goran et al. [64]	EQ-5D	− 0.0215 (0.0043)	888	72.9	51.8	5.1	GR
22	Peters et al. [65]	EQ-5D	− 0.0433 (0.0031)	827	67.0	50.9	6.5	GR
23	Cheng et al. [66]	SF	− 0.0530 (0.0074)	316	60.3	52.8	2.6	HT
24	Fortin et al. [67]	SF	− 0.0101 (0.0038)	238	59.0	71.0	5.3	HT
25	Garrido-Abejar et al. [68]	SF	− 0.0131 (0.0067)	281	82.0	55.0	1.8	HT
26	Lim et al. [69]	SF	− 0.0216 (0.0020)	2645	40.4	62.1	0.4	HT
27	Loza et al. [70]	SF	− 0.0414 (0.0020)	2192	46.1	53.7	2.8	HT
28	Luo et al. [71]	SF	− 0.1039 (0.0007)	75,198	63.5	100	1.5	HT
29	Naveiro-Rilo et al. [72]	SF	− 0.0504 (NA) ^b	369	79.9	55.1	4.5	HT
30	Prazeres and Santiago [73]	SF	− 0.0263 (0.0031)	521	58.2	64.1	4.5	HT
31	Tyack et al. [74]	SF	0.0001 (0.0018)	351	58.8	58.4	7.0	HT
32	Williams and Egede [75]	SF	− 0.0434 (0.0007)	23,789	55 ^d	51.9	1.2	HT

SE standard error, NA not available, GR group regression, HT Hanmer transformation, Pub published, QoL quality of life

^aStandard error estimated via group regression as not reported for published slope

^bStandard error could not be estimated due to missing group sizes

^cStandard error could not be estimated due to missing group variances

^dMedian

borrowing strength can be easily observed when comparing Fig. 1a, b. The pooled effect across all three QoL scales indicates a reduction of 3.84% per additional disease.

Down-weighting evidence derived using the Hanmer transformation had little impact on the results (see Table 2, row 3). Reducing weight by 30% produced slightly less shrinkage towards the overall effect for the SF-6D data, and slightly less reduction in uncertainty.

The frequentist analysis is reported in Table 2, row 4. Outcomes for the EQ-5D and SF-36/12 are very similar to the independent Bayesian analysis. A much smaller uncertainty is observed for the SF-6D analysis, compared with the Bayesian analysis. This is caused by the stronger influence of the prior distribution in the Bayesian case, which was chosen to be very broad. A naïve pooling of all studies results in an overall estimate with much less uncertainty compared with the Bayesian hierarchical model. However,

Table 2 Result summary showing posterior mean and standard deviation of the slope, indicating the change in quality of life with additional disease burden

	SF-6D	EQ-5D	SF-36/12	Overall effect
Independent effects (Bayesian MA)	– 0.0357 (0.160)	– 0.0417 (0.008)	– 0.0362 (0.012)	
Shrunken effects	– 0.0396 (0.029)	– 0.0411 (0.007)	– 0.0377 (0.009)	– 0.0384 (0.089)
Shrunken effects DW70	– 0.0399 (0.028)	– 0.0410 (0.007)	– 0.0376 (0.010)	– 0.0391 (0.095)
Frequentist MA	– 0.0402 (0.016)	– 0.0418 (0.009)	– 0.0363 (0.013)	– 0.0397 (0.005)

Results are shown on instrument level and overall level as appropriate for (1) the independent meta-analyses [independent effects (Bayesian MA)]; (2) the shrunken effects from the three-level model (shrunken effects); (3) results from the model down-weighting evidence derived via the Hanmer transformation by 30% (shrunken effects DW70); and (4) results from a frequentist meta-analysis

MA meta-analysis, DW70 down-weight to 70%

Fig. 1 Posterior distribution of the slope parameter indicating the change in quality of life per additional condition. Results are shown for individual quality-of-life scales for **a** the individual analysis model

and **b** the shrunken effects from the three-level model (not down-weighting evidence), as well as **c** an overall effect on quality of life combining all three scales

naïve pooling does not take into account differences between HRQoL outcomes.

5 Discussion

Evidence evaluating HRQoL in a group of patients relevant for CUA is often heterogeneous. Heterogeneity can be found in the population studied, as well as in the instrument used to collect the outcome. In practice, utility parameters

in CUAs are often informed by a single study, selected to mirror the relevant patient population most appropriately and measuring HRQoL using the requested instrument of the jurisdiction. This may be appropriate in many situations, where alternative sources of evidence are either not available or unsuitable to provide valid information for the parameter. However, there may also be situations where relevant evidence is disregarded, which could reduce decision uncertainty and therefore reduce the potential of taking a suboptimal decision.

Using a case study in MM, we have illustrated how a hierarchical modelling approach can aid in the systematic inclusion of studies evaluating HRQoL using multiple instruments to increase precision in the outcome. We have applied a model, which has previously been used to combine data from studies of different designs [33, 34].

The focus in combining data from different study designs was the overall treatment effect. The hierarchical model structure allowed for design differences, possible adjustments for bias or down-weighting specific designs via variance inflation. The overall aim was to obtain evidence-based estimates of relative treatment effects incorporating the full range of existing evidence across designs. In contrast, the focus of applying a Bayesian hierarchical model to HRQoL studies are the instrument-level estimates. Here, one is interested in using relevant evidence, including evidence on other outcome scales, to strengthen the evidence of individual outcome scales. Rather than analysing each outcome scale individually assuming no correlation, the random effect assumption in the hierarchical model allows reducing uncertainty by borrowing strength across all evidence, while recognising differences between scales.

Individual analyses of outcome measures in this analysis indicated similar effect sizes for all outcomes, as could be seen in the large overlap of posterior distributions in Fig. 1. Hence, heterogeneity between outcomes was relatively small, allowing for a strong effect of borrowing strength. In the hierarchical model, precision was increased for all HRQoL instruments; however, the effect was strongest for SF-6D, where only three studies were included (as opposed to 19 and 10 studies for the other measures). Where little evidence is available for the instrument of interest, a Bayesian hierarchical model such as this one can therefore help to increase precision and thereby reduce decision uncertainty in a CUA. McCullagh et al. have evaluated the expected value of perfect information associated with various parameter categories in the evaluation of cancer drugs in Ireland between 2010 and 2014. Their analysis found uncertainty in utility parameters to be associated with the second highest value of information, highlighting the importance of accurate utility estimates in CUA [76].

There are some limitations to this analysis. The proposed model enabled borrowing strength across HRQoL instruments via a random effects assumption to combine evidence. While the model allows for differences between HRQoL instruments, it assumes a certain similarity modelled using a normal distribution for the random effects. We used a non-informative prior to allow the data to determine the level of heterogeneity. This means that the effect of shrinkage towards an overall mean and reduction in uncertainty will be stronger, where differences between outcome measures are small. Where outcome measures show very different effects, higher heterogeneity will result in less shrinkage and less

increased precision of instrument-level estimates. However, in situations where an individual outcome measure is capable of measuring changes in utility that are not captured by other outcome scales, and hence provides fundamentally different results, there is a risk of diminishing these differences when evidence from the remaining outcomes is too strong. Combining evidence from different utility studies remains a complex issue. Heterogeneity between outcome scales caused by differences in methodological approaches to valuation, as well as fundamental conceptions of health, have shown to lead to major differences in outcomes. The validity of indirect evaluation methods, including mapping, is debated, even though its value to reimbursement decision is clearly acknowledged [77]. The process of valuing preference-based utilities requires complex methodologies and may lead to heterogeneity even within outcome scales [78]. Hence, the selection of outcome scales included in the model remains critical. While the proposed model allows for a systematic inclusion of outcome scales that acknowledges some differences between these, careful consideration of the similarity assumption given by the RE model should guide the selection process.

All included studies were of an observational nature as no randomised controlled studies could be identified. Studies included in this analysis were very heterogeneous. They were conducted over a long time horizon, and across different countries, while populations differed highly in terms of age, sex and number of co-occurring conditions. A particular source of heterogeneity was the definition and assessment of MM. Most authors define MM as the occurrence of two or more diseases, however some define a minimum of three diseases and many do not specify a definition. The number and type of conditions considered in each study differed hugely, ranging from 4 to 147; some studies applied a closed list of diseases to assess morbidity, while others applied an open-end list allowing additional conditions to be added on an individual basis. Additionally, studies varied in their data collection method; some collected information through self-report questionnaires, while others used medical chart records. A detailed discussion and information on individual diseases considered in each study is provided elsewhere [24]. The application of utility estimates to particular patient groups should hence be interpreted with care as additional research on the role of disease patterns and other sources of variation is required.

Our analysis transformed MCS and PCS outcomes to SF-6D estimates using an algorithm incorporating study-level covariate information. The additional uncertainty incurred through the transformation is not automatically incorporated. While the fact that resulting estimates are very similar to those reported in studies measuring SF-6D directly is reassuring, incorporation of evidence without adjustment may still underestimate the uncertainty. A possible solution

is to down-weight evidence obtained through transformation via variance inflation, as has been proposed in the three-level model applied to different study designs, where observational evidence was down-weighted compared with randomised controlled trial evidence [34].

While we aimed at extracting slope parameters that were adjusted for potential confounders, adjustment factors differed between studies, and, on occasion, studies only reported unadjusted outcomes. Our MA model assumes a random effect to combine data from different studies, and thereby allows for differences between studies; however, the analysis still models true effects of each individual study to be randomly scattered around an overall effect.

The outcome measure of interest in this analysis was the linear slope parameter relating HRQoL and increasing number of diseases. This of course makes the assumption of linearity of this relationship. In an MA, we rely on data published, and a linear model was the one most commonly fitted to explore this relationship, which guided our choice. Barra et al. have investigated the comparative fit of several model choices and concluded that additive and multiplicative models perform equally well [45]. On the other hand, the argument of ‘double jeopardy’ provides a theoretical argument against a linear relationship, where, with every additional disease, the ability of being healthy is a further constraint, resulting in smaller and smaller reductions in QoL as the number of diseases increases [79]. Additional studies investigating the relationship between MM and QoL are needed to confirm the optimal modelling approach. Where deemed appropriate, the Bayesian hierarchical model can be adjusted to fit more complex relationships such as quadratic models.

Our analysis measures MM as the number of co-occurring conditions as this was the most widely reported scale in the identified literature. However, the number of diseases does not take into account the severity associated with individual diseases or the stage of disease of the individual subject. Other measures, such as the Charlson comorbidity index, the Cumulative Illness Rating Scale, or the Seattle Index of Comorbidity combine severity and number of diseases and may provide a more distinguished measure for MM.

6 Conclusions

Our analysis has illustrated the use of three-level hierarchical models in the MA of HRQoL data to allow for the inclusion of a large body of evidence. In our case study, we have shown how the analysis of multiple outcomes within one model can reduce uncertainty at the instrument level, which may eventually reduce decision uncertainty in CUA. Our analysis confirms that patients suffer from reduced HRQoL as the number of diseases increases. This effect was observed for all measures of HRQoL investigated.

Author Contributions SSc extracted the data, conducted the statistical analysis, interpreted the results, and wrote the first draft of the manuscript. TM conducted the literature review, extracted and checked the data, and provided detailed feedback on the manuscript. RA provided important input on the application of HRQoL in CUA, and provided detailed feedback on the manuscript. MvA extracted data and provided feedback on the manuscript. SSSt and MZ reviewed and provided important feedback both during the analysis phase and for the manuscript.

Data Availability Statement R code and Winbugs models, together with input data, are available as ESM 1. The PRISMA diagram and the reason for excluding studies are also available online as ESM 2 and 3.

Compliance with Ethical Standards

Funding No funding was received to conduct this study.

Conflict of interest Susanne Schmitz, Tatjana Makovski, Roisin Adams, Marjan van den Akker, Saverio Stranges, and Maurice P. Zeegers have no conflicts of interest to report in respect of this study.

References

1. García-Mochón L, et al. HTA and decision-making processes in Central, Eastern and South Eastern Europe: results from a survey. *Health Policy*. 2019;123(2):182–90.
2. Drummond MF, et al. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press; 2015.
3. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care*. 1992;30(6):473–83.
4. Skevington SM, Lotfy M, O’Connell KA. The World Health Organization’s WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. *Qual life Res*. 2004;13(2):299–310.
5. Oppe M, Devlin NJ, Szende A. *EQ-5D value sets: inventory, comparative review and user guide*. New York: Springer; 2007.
6. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ*. 2002;21(2):271–92.
7. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annu Rev Public Health*. 2000;21(1):587–611.
8. Oppe M, et al. EuroQol protocols for time trade-off valuation of health outcomes. *Pharmacoeconomics*. 2016;34(10):993–1004.
9. Oppe M, et al. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health*. 2014;17(4):445–53.
10. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care*. 2004;42(9):851–9.
11. Tsuchiya A, Brazier J, Roberts J. Comparison of valuation methods used to generate the EQ-5D and the SF-6D value sets. *J Health Econ*. 2006;25(2):334–46.
12. Longworth L, Bryan S. An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Econ*. 2003;12(12):1061–7.
13. Earnshaw J, Lewis G. *NICE guide to the methods of technology appraisal*. New York: Springer; 2008.
14. National Centre for Pharmacoeconomics Ireland. Applicant template for submission of full pharmacoeconomic assessments to the National Centre for Pharmacoeconomics [cited 11 Sep 2019].

- <http://www.ncpe.ie/submission-process/submission-templates/format-of-full-submissions/>.
15. Ara R, et al. The identification, review and synthesis of health state utility values from the literature. *Pharmacoeconomics*. 2017;35(1):43–55.
 16. Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J Clin Nurs*. 2003;12(1):77–84.
 17. Petrou S, Kwon J, Madan JJP. A practical guide to conducting a systematic review and meta-analysis of health state utility values. *Pharmacoeconomics*. 2018;36(9):1043–61.
 18. van den Akker M, Buntinx F, Knottnerus JA. Comorbidity or multimorbidity: what's in a name? A review of literature. *Eur J Gen Pract*. 1996;2(2):65–70.
 19. Fortin M, et al. A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. *Ann Fam Med*. 2012;10(2):142–51.
 20. Agborsangaya CB, et al. Multimorbidity prevalence and patterns across socioeconomic determinants: a cross-sectional survey. *BMC Public Health*. 2012;12(1):201.
 21. Fortin M, et al. Prevalence of multimorbidity among adults seen in family practice. *Ann Fam Med*. 2005;3(3):223–8.
 22. Glynn LG, et al. The prevalence of multimorbidity in primary care and its effect on health care utilization and cost. *Fam Pract*. 2011;28(5):516–23.
 23. Marengoni A, et al. Aging with multimorbidity: a systematic review of the literature. *Ageing Res Rev*. 2011;10(4):430–9.
 24. Makovski TT, et al. Multimorbidity and quality of life: systematic literature review and meta-analysis. *Ageing Res Rev*. 2019;53:100903.
 25. Borenstein M, et al. Introduction to meta-analysis. New York: Wiley; 2011.
 26. Deeks JJ, Higgins JP, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JP, Green S, editors. *Cochrane handbook for systematic reviews of interventions: Cochrane book series*. New York: Wiley; 2008. p. 243–96.
 27. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controll Clin Trials*. 1986;7(3):177–88.
 28. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med*. 2002;21(16):2313–24.
 29. Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105–24.
 30. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21(11):1559–73.
 31. Dias S, et al. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. London: National Institute for Health and Clinical Excellence; 2011.
 32. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res*. 2001;10(4):277–303.
 33. Mesgarpour B, et al. Safety of off-label erythropoiesis stimulating agents in critically ill patients: a meta-analysis. *Intensive care Med*. 2013;39(11):1896–908.
 34. Schmitz S, Adams R, Walsh C. Incorporating data from various trial designs into a mixed treatment comparison model. *Stat Med*. 2013;32(17):2935–49.
 35. Brazier J, et al. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ*. 2004;13(9):873–84.
 36. Bryan S, Longworth L. Measuring health-related utility. *Eur J Health Econ*. 2005;6(3):253–60.
 37. McDonough CM, et al. Comparison of EQ-5D, HUI, and SF-36-derived societal health state values among spine patient outcomes research trial (SPORT) participants. *Qual Life Res*. 2005;14(5):1321–32.
 38. Moher D, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg*. 2010;8(5):336–41.
 39. Core Team R. R: a language and environment for statistical computing. Vienna: R Core Team; 2013.
 40. Hanmer J. Predicting an SF-6D preference-based score using MCS and PCS scores from the SF-12 or SF-36. *Value Health*. 2009;12(6):958–66.
 41. Lunn DJ, et al. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput*. 2000;10(4):325–37.
 42. Sturtz S, Ligges U, Gelman AE. R2WinBUGS: a package for running WinBUGS from R. *J Stat Softw*. 2005;12(3).
 43. Prevost TC, Abrams KR, Jones DR. Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. *Stat Med*. 2000;19(24):3359–76.
 44. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3):1–48.
 45. Barra M, et al. Examining the relationship between health-related quality of life and increasing numbers of diagnoses. *Qual Life Res*. 2015;24(12):2823–32.
 46. Rabadi MH, Vincent AS. Health status profile and health-related quality of life of veterans attending an out-patient clinic. *Med Sci Monit*. 2013;19:386–92.
 47. Hanmer J, et al. Three methods tested to model SF-6D health utilities for health states involving comorbidity/co-occurring conditions. *J Clin Epidemiol*. 2010;63(3):331–41.
 48. Lawson KD, et al. Double trouble: the impact of multimorbidity and deprivation on preference-weighted health related quality of life a cross sectional analysis of the Scottish Health Survey. *Int J Equity Health*. 2013;12:67.
 49. Agborsangaya CB, et al. Comparing the EQ-5D 3L and 5L: measurement properties and association with chronic conditions and multimorbidity in the general population. *Health Qual Life Outcomes*. 2014;12(1):74.
 50. Agborsangaya CB, et al. Health-related quality of life and health-care utilization in multimorbidity: results of a cross-sectional survey. *Qual Life Res*. 2013;22(4):791–9.
 51. Alonso-Morán E, et al. Health-related quality of life and multimorbidity in community-dwelling telecare-assisted elders in the Basque Country. *Eur J Intern Med*. 2015;26(3):169–75.
 52. Ferrer A, et al. Health-related quality of life in nonagenarians: gender, functional status and nutritional risk as associated factors. *Med Clin (Barc)*. 2010;134(7):303–6 (in Spanish).
 53. Gerber AM, et al. A cohort study of elderly people in Bloemfontein, South Africa, to determine health-related quality of life and functional abilities. *S Afr Med J*. 2016;106(3):298–301.
 54. Heyworth IT, et al. How do common chronic conditions affect health-related quality of life? *Br J Gen Pract*. 2009;59(568):e353–8.
 55. Hodek JM, Ruhe A, Greiner W. Multimorbidity and health-related quality of life among elderly persons. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2009;52(12):1188–201 (in German).
 56. Kim KI, Lee JH, Kim CH. Impaired health-related quality of life in elderly women is associated with multimorbidity: results from the Korean national health and nutrition examination survey. *Gender Med*. 2012;9(5):309–18.
 57. Lang K, et al. The impact of multimorbidity on quality of life among midlife women: findings from a US nationally representative survey. *J Womens Health (Larchmt)*. 2015;24(5):374–83.
 58. Li J, et al. Patterns of multimorbidity and their association with health outcomes within Yorkshire, England: baseline results from the Yorkshire Health Study. *BMC Public Health*. 2016;16:649.
 59. Mujica-Mota RE, et al. Common patterns of morbidity and multi-morbidity and their impact on health-related

- quality of life: evidence from a national survey. *Qual Life Res.* 2015;24(4):909–18.
60. Quah JHM, et al. Health-related quality of life of older Asian patients with multimorbidity in primary care in a developed nation. *Geriatr Gerontol Int.* 2017;17(10):1429–37.
 61. Sakthong P, et al. Psychometric properties of the EQ-5D-5L in Thai patients with chronic diseases. *Qual Life Res.* 2015;24(12):3015–22.
 62. Sullivan PW, Ghushchyan VH, Bayliss EA. The impact of co-morbidity burden on preference-based health-related quality of life in the United States. *Pharmacoeconomics.* 2012;30(5):431–42.
 63. Vogel I, et al. The impact of perceived social support and sense of coherence on health-related quality of life in multimorbid primary care patients. *Chronic Illn.* 2012;8(4):296–307.
 64. N’Goran AA, et al. Comparing the self-perceived quality of life of multimorbid patients and the general population using the EQ-5D-3L. *PLoS One.* 2017;12(12):e0188499.
 65. Peters M, et al. Quality of life and burden of morbidity in primary care users with multimorbidity. *Patient Relat Outcome Meas.* 2018;9:103.
 66. Cheng L, et al. Health related quality of life in pregeriatric patients with chronic diseases at urban, public supported clinics. *Health Qual Life Outcomes.* 2003;1:63.
 67. Fortin M, et al. Relationship between multimorbidity and health-related quality of life of patients in primary care. *Qual Life Res.* 2006;15(1):83–91.
 68. Garrido-Abejar M, et al. Factors associated with health-related quality of life in the institutionalised elderly: differences between men and women. *Enferm Clin.* 2012;22(1):27–34 (in Spanish).
 69. Lim L, Jin AZ, Ng TP. Anxiety and depression, chronic physical conditions, and quality of life in an urban population sample study. *Soc Psychiatry Psychiatr Epidemiol.* 2012;47(7):1047–53.
 70. Loza E, et al. Multimorbidity: prevalence, effect on quality of life and daily functioning, and variation of this effect when one condition is a rheumatic disease. *Semin Arthritis Rheum.* 2009;38(4):312–9.
 71. Luo J, et al. Newly developed chronic conditions and changes in health-related quality of life in postmenopausal women. *J Am Geriatr Soc.* 2015;63(11):2349–57.
 72. Naveiro-Rilo JC, et al. Quality of life in the elderly on polymedication and with multiple morbidities. *Rev Esp Geriatr Gerontol.* 2014;49(4):158–64 (in Spanish).
 73. Prazeres F, Santiago L. Relationship between health-related quality of life, perceived family support and unmet health needs in adult patients with multimorbidity attending primary care in Portugal: a multicentre cross-sectional study. *Health Qual Life Outcomes.* 2016;14(1):156.
 74. Tyack Z, et al. Predictors of health-related quality of life in people with a complex chronic disease including multimorbidity: a longitudinal cohort study. *Qual Life Res.* 2016;25(10):2579–92.
 75. Williams JS, Egede LE. The association between multimorbidity and quality of life, health status and functional disability. *Am J Med Sci.* 2016;352(1):45–52.
 76. McCullagh L, et al. Examining the feasibility and utility of estimating partial expected value of perfect information (via a nonparametric approach) as part of the reimbursement decision-making process in Ireland: application to drugs for cancer. *Pharmacoeconomics.* 2017;35(11):1177–85.
 77. McCabe C, Edlin R, Hall P. Navigating time and uncertainty in health technology appraisal: would a map help? *Pharmacoeconomics.* 2013;31(9):731–7.
 78. Dolan P. Modeling valuations for EuroQol health states. *Med Care.* 1997;35(11):1095–108.
 79. Harris J. Double jeopardy and the veil of ignorance: a reply. *J Med Ethics.* 1995;21(3):151–7.

Affiliations

Susanne Schmitz¹ · Tatjana T. Makovski^{2,3,4} · Roisin Adams⁵ · Marjan van den Akker^{3,6,7} · Saverio Stranges^{2,8,9} · Maurice P. Zeegers⁴

¹ Competence Center for Methodology and Statistics, Department of Population Health, Luxembourg Institute of Health, 1 A-B, rue Thomas Edison, 1445 Strassen, Luxembourg

² Epidemiology and Public Health Research Unit, Department of Population Health, Luxembourg Institute of Health, Strassen, Luxembourg

³ Department of Family Medicine, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands

⁴ Chairgroup of Complex Genetics and Epidemiology, Nutrition and Metabolism in Translational Research (NUTRIM), Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands

⁵ National Centre for Pharmacoeconomics, St James’s Hospital, Dublin, Ireland

⁶ Institute of General Practice, Johann Wolfgang Goethe University, Frankfurt am Main, Germany

⁷ Academic Centre of General Practice, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

⁸ Department of Epidemiology and Biostatistics, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada

⁹ Department of Family Medicine, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada