PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A protocol for a systematic review and meta-analysis of minimal important differences for generic multi-attribute utility instruments
AUTHORS	Henson, Glen; Taylor, Bruce; van der Mei, Ingrid; Claflin, Suzi; Simpson-Yap, Steve; Palmer, Andrew; Xia, Qing; Antony, Benny; Singh, Ambrish; Campbell, Julie

VERSION 1 – REVIEW

REVIEWER	Are Hugo Pripp Oslo universitetssykehus Ulleval, Oslo Centre for Biostatistics & Epidemiology
REVIEW RETURNED	04-May-2022

GENERAL COMMENTS	Statistical review: The statistical analysis plan in this protocol article seems adequate. Please add some information about the type of statistical software you plan to use for the meta-analysis. I think that "Generic multi-attribute utility instruments" could be somewhat new to many readers. Maybe you could add an illustration or some examples to explain it and its use? Please also avoid too many abbreviations in the text to ease readability.

REVIEWER	Joseph Kwon University of Oxford Nuffield Department of Primary Care Health Sciences
REVIEW RETURNED	01-Jun-2022

GENERAL COMMENTS	Thank you for this submission; it was enlightening to read it. I hope the comments and suggestions herein can help improve the manuscript. Introduction 1. “These algorithms generate health state utilities (HSUs), which are ordinal rankings of health-related quality of life” (p. 4, lines 7-9). Perhaps it may be worth mentioning the cardinal property of HSUs: i.e., the magnitude of difference between HSUs of two health states carries evaluative significance. 2. Please state upfront that MAUIs can be generic or disease-specific in Section 1.1, not later in Section 1.4. 3. Table 1: HUI3 has 15 items if self-administered and 40 if interview-administered; please correct. 4. Please add ‘generic’ to p. 7, line 13 to read “MIDs for generic MAUIs”. Methods

	1. Please use the latest PRISMA guideline for reporting of systematic reviews: Page et al (2021) – The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. 2. Please elaborate on how the CHEERS checklist for the reporting of economic evaluations will be used to "determine the suitability of studies, meeting inclusion criteria, for incorporation into the systematic review" (p. 8, lines 22-24). 3. Is the ROBIS tool used to assess the risk of bias of systematic review because the COSMIN guideline for systematic reviews of patient-reported outcome measures (Prinsen et al., 2018) does not contain adequate risk of bias assessment criteria? If so, please state this in Section 2.4.

VERSION 1 – AUTHOR RESPONSE

Reviewer One's Comments

Comment 5 – Please add some information about the type of statistical software you plan to use for the meta-analysis.

Response – The study has been amended to say that Stata 17 will be used in numerical analyses. See marked changes below.

2.11 Meta-analysis

Provided that sufficient data is extracted from studies meeting the inclusion criteria (specifics regarding what comprises sufficient data are currently unknown), meta-analyses and meta-regressions will be performed using Stata 17 (Stata Corp, 2022).

Comment 5 – I think that "Generic multi-attribute utility instruments" could be somewhat new to many readers. Maybe you could add an illustration or some examples to explain it and its use?

Response – Revised section 1.1 "Multi-Attribute utility instruments" to present a more detailed and clearer explanation of multi-attribute utility instruments, their function, and their uses. In addition, a new figure was constructed and captioned, as suggested, to provide further clarity. See marked changes and the new figure below.

1.1 Multi-attribute utility instruments

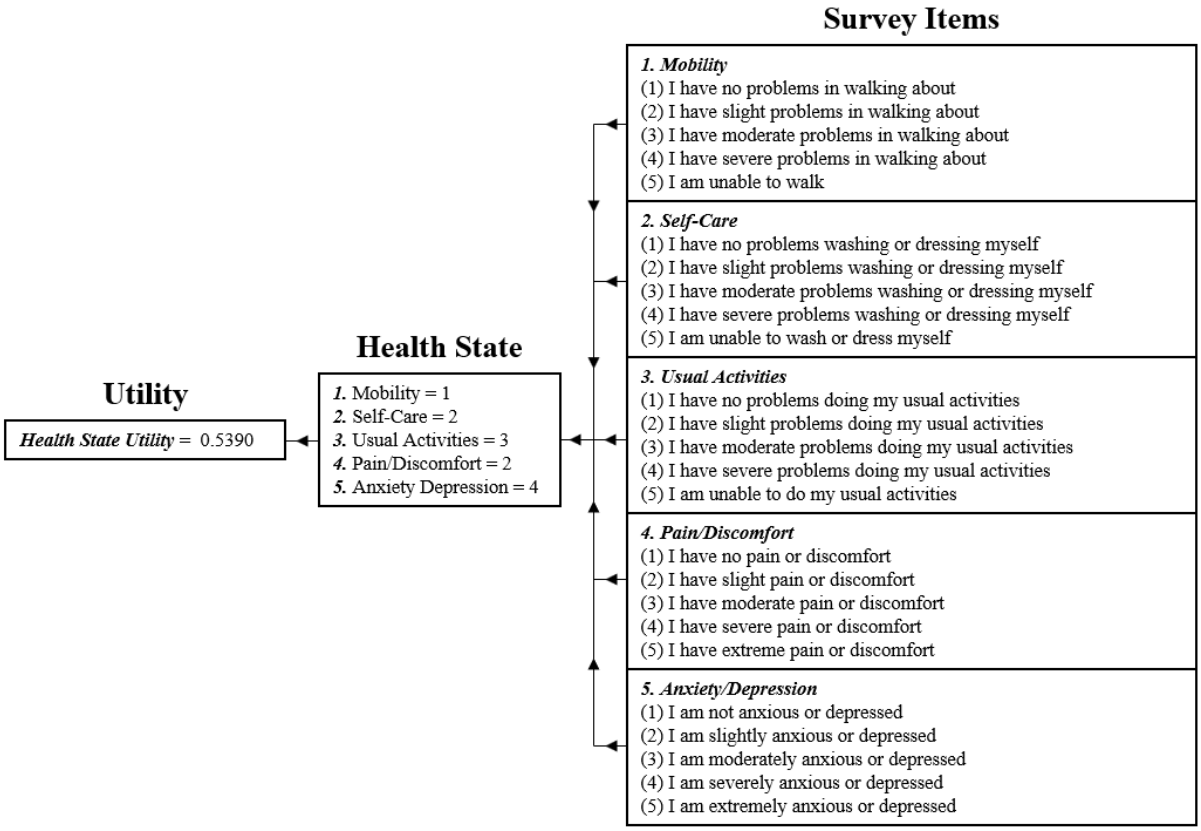
Multi-attribute utility instruments can be generic and adopted for use with any study population/sample, or be disease- or symptom-specific. Multi-Attribute utility instruments operate by eliciting health states, which are profiles of overall health-related quality of life across several dimensions of health. Multi-attribute utility instruments health states are based on arrays of patient-reported outcomes, obtained through multi-attribute utility instrument-specific surveys [1].

Multi-Attribute utility instruments surveys function by posing questions about several physical and psychosocial dimensions of health [2]. These questions require respondents to rank their dimensional health [2]. Uniquely, the Assessment Quality of Life – 8 Dimensions (AQoL-8D)[3]

generic multi-attribute utility instrument coalesces dimensional scores into super-dimensional scores, which provides measures of overall physical and mental health. Other common, generic multi-attribute utility instruments include the European Quality of Life – 5 Dimensions – 5 Levels (EQ-5D-5L) [4], Quality of Wellbeing (QWB) [5], Short Form – 6 Dimensions Version 1 (SF-6Dv1) [6], and Health Utilities Index Version 3 (HUI3) [7], which all vary in size and the health dimensions they assess. See *Table 1* for a list of common, generic multi-attribute utility instruments, the dimensions of health they analyse, and the number of items (questions) in each.

Each health state, generatable by a multi-attribute utility instrument via its survey, has an associated health state utility, which is a discrete, ordinal ranking of health-related quality of life [8]. These health state utilities are assigned to health states using a variety of experimental economics techniques including standard gambles, visual-analogue scales, discrete choice experiments, and time trade-offs [9]. Health state utilities are best defined as representing the position of a person's health state on a death (0) to full health (1) continuum, relative to the positions of all other possible health states. The representation of health state utilities as a pseudo-continuous measure is facilitated by the large number of health states identifiable by multi-attribute utility instruments. For example, the AQoL-8D can generate 2.4×10^{23} discrete health states [3]. Health state utilities are frequently applied in cost-utility analyses (a type of full, health economic analysis used to evaluate medical interventions), clinical assessments, and evaluations of patient-reported outcomes [1 4]. In *Figure 1* the function of the EQ-5D-5L is presented to exemplify the operation of a generic multi-attribute utility instrument.

Figure 1 – Operation of the EQ-5D-5L multi-attribute utility instrument



6.1 Figure 1 Caption

This figure illustrates the function of the EQ-5D-5L multi-attribute utility instrument. The first element of the process involves obtaining participant responses to the relevant multi-attribute utility instrument survey. In the case of the EQ-5D-5L, participants are required to select one of five ranks for each of the five survey items (questions). These responses are then collated and used to produce a profile of participant health, known as a health state. Finally, the health state utility associated with the participant's health state is retrieved, usually via an algorithm.

Comment 6 – Please also avoid too many abbreviations in the text to ease readability.

> Response – We have now removed most of the abbreviations. The only remaining frequently used acronym in the manuscript is minimal important difference (MID). All other abbreviations were expanded into their full-text forms. For example, multi-attribute utility instrument (MAUI) and health state utility (HSU) no longer appear as acronyms. Note that some database and guidelines/quality appraisal framework names still appear as acronyms. However, these are always spelled out in adjacent text.

Reviewer Two's Comments

Comment 7 - "These algorithms generate health state utilities (HSUs), which are ordinal rankings of health-related quality of life" (p. 4, lines 7-9). Perhaps it may be worth mentioning the cardinal

property of HSUs: i.e., the magnitude of difference between HSUs of two health states carries evaluative significance.

Response – Thank you for this suggestion. I have added the following line to the manuscript.

This attribute also allows the magnitude of difference between health state utilities to bear comparative significance, adding an element of cardinality to an otherwise ordinal measure.

As seen above, I have used the word 'comparative' rather than 'evaluative'. Respectfully, I would suggest that this is more appropriate. This is because the magnitude of difference between HSUs is representative of the distance between two health states on a continuum (or alternatively, of the number of health states which are positioned, in order of preference, between the two health states of interest). This speaks to the difficultly in evaluating changes in HSU in the absence of MIDs or comparable changes in HSU (which could be, for example, the change in HSU attributable to the implementation of the current best practice).

Comment 8 - Please state upfront that MAUIs can be generic or disease-specific in Section 1.1, not later in Section 1.4.

Response – Section 1.1 was updated in line with this recommendation. See marked changes below.

1.1 Multi-attribute utility instruments

Multi-attribute utility instruments can be generic (usable with any study cohort) or disease or symptom-specific.

Comment 9 – Table 1: HUI3 has 15 items if self-administered and 40 if interview-administered; please correct.

Response – Correction made. The table now reads "HUI3 Self-Administered … 15 items".

Comment 10 - Please add 'generic' to p7, line 13 to read "MIDs for generic MAUIs".

Response – Correction made.

Comment 11 – Please use the latest PRISMA guideline for reporting of systematic reviews: Page et al (2021) – The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.

Response – References updated appropriately. The new reference is reproduced below.

31. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an

updated guideline for reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097. BMJ. 2021;372.doi:10.1371/journal.pmed.10000971136/bmj.n71.

Comment 12 – Please elaborate on how the CHEERS checklist for the reporting of economic evaluations will be used to "determine the suitability of studies, meeting inclusion criteria, for incorporation into the systematic review" (p. 8, lines 22-24).

Response: The CHEERS checklist was inappropriate to "determine the suitability of studies, meeting inclusion criteria, for incorporation into the systematic review". Sections 2.2 and 2.3 were overhauled accordingly. See marked changes below.

2.2 Validated guidelines: protocol and systematic review

This protocol has been developed according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses Protocols guidelines (PRISMA-P) [30]. The proposed systematic review will be performed and reported in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses guidelines (PRISMA) [31]. The review will also adhere to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) [32] checklist and the Professional Society for Health Economics and Outcomes Research (ISPOR) good research practices task force report regarding health state utilities in clinical studies [33].

2.3 Validated guidelines: quality appraisal and risk of bias assessment for included studies

The COSMIN methodology for patient-reported outcome measures assessment checklist will be adapted and applied to evaluate the quality of papers considered for inclusion in the study, as well as their associated risk of bias [34-37]. Any studies found to be at high risk of bias will be weighted in meta-analysis, to reduce their impact on review conclusions. Additionally, references from included papers will be screened for relevant articles to identify potential omissions in the systematic review, thereby ensuring quality through completeness.

Comment 13 – Is the ROBIS tool used to assess the risk of bias of systematic review because the COSMIN guideline for systematic reviews of patient-reported outcome measures (Prinsen et al., 2018) does not contain adequate risk of bias assessment criteria? If so, please state this in Section 2.4.

Response – The COSMIN methodology can be applied in risk of bias assessments for individual studies, whereas the ROBIS instrument is specifically designed to evaluate the risk of bias in a systematic review's body of evidence. Adjustments were made to convey this. In addition, the GRADE framework was included in the protocol, as previously there was no validated framework included which could be used to evaluate the overall strength and certainty of the systematic review's body of evidence. See marked changes below.

2.4 Validated guidelines: evidence appraisal and risk of bias assessment for the systematic review

To assess the overall risk of bias in the systematic review's body of evidence, the Risk of BIas assessment tool for Systematic reviews (ROBIS) was selected [38]. The ROBIS tool has several domains under which bias may be judged: study eligibility criteria (did the study adhere to predefined eligibility criteria), identification and selection of studies (was every effort made to collect the maximum number of eligible papers), data collection and study appraisal (was potential bias in individual studies assessed and all pertinent data collected), and synthesis and findings (was all available data synthesised appropriately and any potential bias in results made transparent) [38]. In addition, to evaluate the overall certainty and strength of the body of evidence generated by the systematic review, the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework will be implemented [39].

VERSION 2 – REVIEW

REVIEWER	Joseph Kwon University of Oxford Nuffield Department of Primary Care Health Sciences
REVIEW RETURNED	10-Sep-2022

GENERAL COMMENTS	The authors have adequately addressed the comments and suggested revisions on the manuscript.