BMJ Open

Protocol for a systematic review and meta-analysis of minimal important differences for generic multiattribute utility instruments

Glen James Henson ,[1] Bruce V Taylor ,[2] Ingrid van der Mei ,[2] Suzi B Claflin ,[2] Steve Simpson-Yap ,[3] Andrew J Palmer ,[1] Qing Xia ,[1] Benny Antony ,[2] Ambrish Singh ,[2] Julie A Campbell [1]

To cite: Henson GJ, Taylor BV, van der Mei I, *et al*. Protocol for a systematic review and meta-analysis of minimal important differences for generic multiattribute utility instruments. *BMJ Open* 2022;**12**:e062703. doi:10.1136/bmjopen-2022-062703

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-062703).

Received 21 March 2022
Accepted 14 September 2022

[1]Health Economics, Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia
[2]Menzies Insitute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia
[3]School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia

Correspondence to
Dr Julie A Campbell;
Julie.Campbell@utas.edu.au

ABSTRACT

Introduction Generic multiattribute utility instruments (MAUIs) are efficient tools for determining and enumerating health-related quality of life. MAUIs accomplish this by generating health state utilities (HSUs) via algorithms. Minimal important differences (MIDs) assist with the interpretation of HSUs by estimating minimum changes that are clinically significant. The overall goal of the proposed systematic review and meta-analysis is the development of comprehensive guidelines for MID estimation.

Methods and analysis This protocol defines a systematic review and meta-analysis of MIDs for generic MAUIs. The proposed research will involve a comprehensive investigation of 10 databases (EconLit, IDEAs database, INAHTA database, Medline, PsycINFO, Embase, Emcare, JBIEBP and CINAHL) from 1 June 2022 to 7 June 2022, and will be performed and reported in accordance with several validated guidelines, principally the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. The quality of papers, considered for inclusion in the review, will be appraised using the COnsensus-based Standards for the selection of health Measurement INstruments, inter alia.
Narrative analysis will involve identifying the characteristics of MIDs including methods of calculation, sources of heterogeneity, and validation. Meta-analysis will also be conducted. The descriptive element of meta-analysis will involve the generation of I^2 statistics and Galbraith plots of MID heterogeneity. Together with narrative analysis, this will allow sources of MID heterogeniety to be identified. A multilevel mixed model, estimated via restricted maximum likelihood estimation, will be constructed for the purposes of meta-regression. Meta-regression will attempt to enumerate the effects of sources of heterogeneity on MID estimates. Meta-analysis will be concluded with pooling of MIDs via a linear random-effects model.

Ethics and dissemination Ethics approval is not required for this review, as it will aggregate data from published literature. Methods of dissemination will include publication in a peer-reviewed journal, as well as presentation at conferences and seminars.

PROSPERO registration number CRD42021261821.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The systematic review will investigate ten databases (both biomedical and economic) and apply a broad range of search terms, both of which will minimise the risk of study omission.
⇒ Restricted maximum likelihood estimation (REML) was chosen for meta-regression to allow for variability in fixed-effects estimates and degrees of freedom consumption.
⇒ Use of REML will permit superior statistical inference compared with generic maximum likelihood estimation.
⇒ A comprehensive suite of validated guidelines is to be adopted in the systematic review to ensure study quality and limit the potential for bias.
⇒ Due to a lack of consistent terminology, relevant articles may be missed if they have paraphrased 'minimal important difference' in an unusual way which is not capture by the systematic review's search strategy.

INTRODUCTION

The following is a protocol for a systematic review and meta-analysis of minimal important differences (MIDs) for generic multiattribute utility instruments.

Multiattribute utility instruments

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

Multiattribute utility instruments surveys function by posing questions about several

Table 1 Health dimensions assessed by eight multiattribute utility instruments, and the number of items in each

Instrument name	Health dimensions assessed		No of items
EQ-5D-5L[4]	► Mobility ► Self-care ► Usual activities	► Pain/discomfort ► Anxiety/depression	5
AQoL-8D[3]	► Independent Living ► Senses ► Pain ► Mental health	1. Happiness 2. Self-Worth 3. Coping 4. Relationships	35
HUI3 (self-administered) [7]	► Vision ► Hearing ► Speech ► Ambulation	► Dexterity ► Emotion ► Cognition ► Pain	15
QWB[9]	► Chronic symptoms ► Acute symptoms ► Mental health	► Mobility ► Usual Activity ► Physical Activity	74
15-D [40]	► Breathing ► Mental function ► Speech (communication) ► Vision ► Mobility ► Usual activities ► Vitality ► Hearing	► Eating ► Elimination ► Sleeping ► Distress ► Discomfort and symptoms ► Sexual Activity ► Depression	15
SF-6Dv1 [41]	► Physical function ► Role limitation ► Social function	► Bodily pain ► Mental health ► Vitality	6
EQ-5D-5L Psychosocial[42]	► Mobility ► Self-care ► Usual activities ► Pain/discomfort ► Anxiety/depression	► Vitality ► Sleep ► Social relationships ► Community connectedness	9
PROPr Scoring System for the PROMIS [43]	► Cognitive function ► Depression ► Fatigue ► Pain Interference	► Physical function ► Sleep disturbance ► Social roles and activities	Variable

AQoL-8D, Assessment Quality of Life-8 Dimensions; 15-D, 15-Dimension; EQ-5D-5L, European Quality of Life-5 Dimensions-5 Levels; HUI3, Health Utilities Index Version 3; PROMIS, Patient-Reported Outcome Measurement Information System; PROPr, PROMIS Preference; QWB, Quality of Wellbeing; SF-6Dv1, Short Form-6 Dimensions Version 1.

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

Each health state, generatable by a multiattribute 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 pseudocontinuous measure is facilitated by the large number of health states identifiable by multiattribute utility instruments. For example, the AQoL-8D can generate 2.4×10^{23} discrete health states.[3] 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. Health state utilities are frequently applied in cost-utility analyses (a type of comprehensive 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

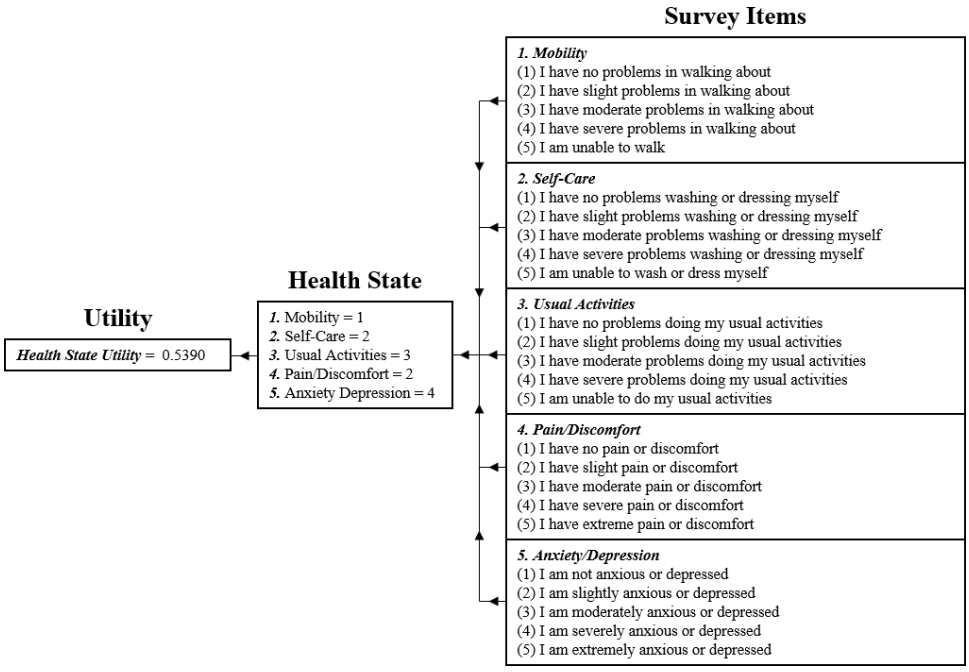

Figure 1 This figure illustrates the function of the EQ-5D-5L multiattribute utility instrument. The first element of the process involves obtaining participant responses to the relevant multiattribute utility instrument survey. In the case of the EQ-5D-5L, participants are required to select one of five ranks for each of the five survey items. 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. EQ-5D-5L, European Quality of Life-5 Dimensions-5 Levels.

EQ-5D-5L is presented to exemplify the operation of a generic multiattribute utility instrument.

Minimal important differences

Although variations in health-related quality of life can be measured using multiattribute utility instruments, these instruments provide no evaluation of what constitutes a clinically significant/meaningful change. Therefore, MIDs are required.[10] These values are the smallest change in health state utility that is statistically significant and represents a meaningful adjustment to patient health-related quality of life.[11] MIDs can lack robustness across multiattribute utility instruments and populations.[12–14]

MID calculation methods

Major methods of MID estimation are described as distribution-based and anchor-based.[10] Distribution-based methods rely on statistical techniques to develop MIDs. An example of such a method is Cohen's effect sizes.[15] Cohen's effect sizes are calculated as $ES = (M_2 - M_1)/S_1$.[15] In this equation, M_1 is the average baseline health state utility for a sample of participants. M_2 is a health state utility greater than the average baseline health state utility, which represents, comparatively, a superior health state. S_1 is the standard error (SE) of the mean, baseline health state utility. Using a classification scale, the output of the equation can be used to classify a change in health state utility as large (not an MID) or small (possibly an MID).[16] Another distribution-based method involves using a fraction of the SE of the mean change in HSU as a MID.[3]

Anchor-based methods can be subdivided into external and internal anchors. External anchors can involve respondents being separately questioned, following multiattribute utility instrument implementation, regarding whether changes in their health state utility represent meaningful changes in their health.[10–17] They can also involve the use of clinical markers to validate the materiality of variations in health state utility. Contrastingly, internal anchors are instrument defined. They are derived as the difference in attributable health state utilities between two minimally different health states, which are thought to be clinically distinct.[11]

Other methods of MID calculation include using legacy anchors, triangulation, and the Delphi method. Legacy anchors are MIDs sourced from previous work and are either reapplied to a new study or used to benchmark new MIDs.[18] Triangulation involves the use of both distribution and anchor-based methods to generate a single MID.[19] MID triangulation is intended to provide increased internal validity to MID estimates.[19] Lastly, the Delphi method involves establishing MIDs by consensus.

Gaps in the literature

No study has been conducted which is a specific and systematic review and meta-analysis of MIDs for generic multiattribute utility instruments. Due to this evidence gap, there are also no guidelines regarding MID estimation for generic multiattribute utility instruments which are validated by a systematic review and meta-analysis. Existing literature has either reviewed MIDs for

multiattribute utility instruments in conjunction with MIDs for disease or symptom-specific instruments[20–23] or focused on MIDs relevant to a particular intervention or disorder.[24 25] Studies applicable to the former category have often been limited in scope, searching few databases.[20 22] Other such studies had different aims than guideline construction, such as highlighting research gaps through systematic review[21] or establishing an MID repository.[23]

Research questions

The proposed systematic review and meta-analysis will address the following research questions regarding MIDs for generic multiattribute utility instruments:

1. How were MIDs calculated?
   - Which methods were applied?
   - Which methods were most commonly used?
   - Were some methods novel and if so in what way?
   - Did different calculation methods produce significantly different MIDs?
2. For what multiattribute utility instruments and diseases were MIDs calculated?
   - Were MIDs consistent across multiattribute utility instruments and diseases?
   - Is variation present in MIDs across iterations using the same, similar, and different study cohorts?
3. Are applied methods of MID estimation theoretically and empirically sound?
   - Were there any mathematical errors or controversial innovations?
   - Were the methods validated?
4. How were MIDs evaluated?
   - What, if any, guidelines were used to evaluate MIDs and were these guidelines validated?
   - What was the result of MID evaluations?
5. What variables, if any, contribute systematically to heterogeneity in MID estimates?
   - Can regression-based evidence be acquired to support relevant associations?
   - If influential variables are controlled for, do MID estimates converge?
   - What level of unexplained heterogeneity exists?
6. Can existing MIDs be applied to new research and under what circumstances?

Aim and rationale

The review aims to generate complete and nuanced guidelines for MIDs for generic multiattribute utility instruments, validated by a systematic review and meta-analysis. Specifically, these guidelines will inform researchers regarding appropriate methods of MID estimation, provide benchmarks against which MIDs may be compared, and expound on potential sources of heterogeneity. Regarding the latter, this will assist researchers in determining the applicability of existing MIDs to new studies and allow benchmark MIDs to have greater comparability to a wider range of MIDs.

METHODS AND ANALYSIS

Patient and public involvement

The was no public or patient involvement, due to the proposed study being a systematic review.

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).[26] The proposed systematic review will be performed and reported in accordance with the PRISMA guidelines.[27] The review will also adhere to the Consolidated Health Economic Evaluation Reporting Standards[28] checklist and the Professional Society for Health Economics and Outcomes Research good research practices task force report regarding health state utilities in clinical studies.[29]

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

The COnsensus-based Standards for the selection of health Measurement INstruments (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.[30–33] Any studies found to be at high risk of bias will be weighted in meta-analysis, to reduce their impact on review results. In addition, references from included papers will be screened for relevant articles to identify potential omissions in the systematic review, thereby ensuring quality through completeness.

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.[34] 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).[34] 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.[35]

Search methodology

A prestudy, preliminary search for relevant papers was conducted using the PubMed database. This permitted collection of keywords appropriate for use in electronic database searches. A professional librarian was enlisted to assist with this task. Collected terms were grouped based on synonymity, as shown in figure 2.

The search strategy selected requires one word or phrase from each of the 'MID' divisions and a phrase or

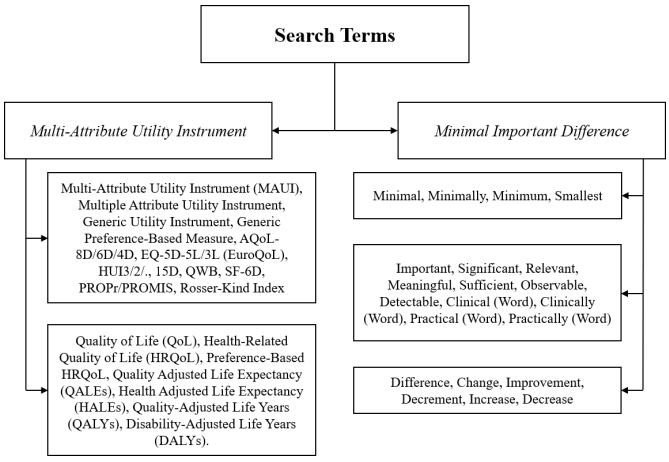

Figure 2 Synonymic groupings of search words associated with 'minimal important difference' are divided by which element of the phrase they are interchangeable with. From top to bottom, the words are associated with 'minimal', 'important' and 'difference'. In addition, '(word)' indicates that singular words from the same category should be added. For example, 'clinically' would become 'clinically important' and 'clinically significant'. Words associated with 'multiattribute utility instrument' are divided into (top) instrument names (generic and specific) and (bottom) outcome measures associated with multiattribute utility instruments. AQoL-8D, Assessment Quality of Life-8 Dimensions; EQ-5D-5L, European Quality of Life-5 Dimensions-5 Levels; QWB, Quality of Wellbeing; SF-6D, Short Form-6 Dimensions.

name from either 'multiattribute utility instrument' division to be present in an article's title and/or abstract for that paper to be considered for inclusion. In addition, search terms will be applied as pluralised (hyphenated) and singular (non-hyphenated) variants. Relevant acronyms are to be applied in searches, as well as their respective expansions. Note that many phrases synonymous with the technical term (MID) are present in the search strategy due to the heterogeneity of their usage and the lack of a firmly established nomenclature.[36] See online supplemental appendix for the precise search strategy used in all database searches.

Both economic and biomedical electronic databases will be searched in this review, from 1 June 2022 to 7 June 2022. Economic databases to be investigated are the American Economic Association database (EconLit) via EBSCO, the IDEAs database by Research Papers in Economics (RePEc), and the International Health Technology Assessment Database (INAHTA). Biomedical databases that are to be examined include Medline, via PubMed and Ovid; PsycINFO, Embase, Emcare, and the Joanna Briggs Institute Evidence-Based Practice (JBIEBP) database via Ovid; and the Cumulative Index to Nursing and Allied Health Literature (CINAHL), via EBSCO. In addition, we will also search Health Business Elite via EBSCO, and google scholar will be utilised to maximise the completeness of the review.

Inclusion criteria

This review will include English language papers that incorporate MIDs for generic multiattribute utility instruments that generate health state utilities. Studies with various response rates, sample sizes and MID calculation techniques will be included, without qualification, to ensure comprehensiveness. No study conducted before 1989 will be considered, as MIDs were introduced into the literature in that year.[37] Furthermore, only original, published studies will be included; editorials, commentaries, protocols, reviews, unpublished works and meta-analyses are to be excluded. Case, in vitro, and animal studies will also be excluded.

Study screening

The first author (GJH) will collect all articles found using the search strategy. Duplicates will be eliminated, and abstracts sorted, using the Covidence program. GJH and JAC will independently screen accumulated papers through analysis of titles and abstracts, excluding those not meeting the inclusion criteria (detailed in section 2.1). The second round of screening (conducted by GJH and JAC) will examine the full text of the remaining articles, excluding articles that fail to satisfy the inclusion criteria, and determining which articles contain sufficient information to be included in meta-analyses. Where disagreements occur during screening, coauthors will be invited to mediate.

Data extraction

Completeness and quality of data extraction will be controlled using a data extraction form. Adherence to this form will be validated by JAC. Where data are not present in a paper, authors will be contacted. The following will be extracted from included studies:

1. Participant characteristics: age, socioeconomic status, sex, education, the urbanity of residences, health insurance coverage, number of participants, diseases and comorbidities, exposure to socialised medicine, countries of residence, response rate, attrition rate, and medication usage.
2. Publication attributes: first and last author, date, journal, country of origin, type of study, quality, risk of bias, and adherence to validated guidelines.
3. Mathematical features: instrument(s) involved, methods of MID calculation, and approach to MID evaluation.
4. Details of sample selection: exclusion criteria, inclusion criteria, and details of participant recruitment method.
5. Results: MID values, MID SEs, and MID robustness.
6. Key discussions: comparisons to the literature, strengths and limitations, and self and peer appraisals of study MID values.

Note that data will be extracted in a qualitative (as opposed to quantitative) form where necessary.

Data management

As noted in section 2.6, extracted abstracts will be sorted, and duplicates removed, using Covidence. After screening, accumulated data will be stored by the first

author (GJH) in Excel spreadsheets and saved on both an institutional cloud and a personal hard drive. The senior author (JAC) will also maintain digital a copy to further ensure data is restorable.

Narrative analysis

Narrative analysis will comprise collation and review of extracted data. For example, methods of MID estimation, frequency of method usage, and context of application will be synthesised into guidelines informing MID application, during this phase. Similar undertakings will occur for other data which does not require further, mathematical analysis. Narrative analysis will also include quality and risk of bias appraisals for included papers.

Meta-analysis

Provided that sufficient data are 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 V.17 (StataCorp, 2022). Descriptive meta-analysis will consist of generating and analysing summary statistics pertaining to MID heterogeneity (including I^2 statistics and Galbraith plots) and undertaking subgroup analysis using stratification. Subgroups will consist of MIDs estimated for specific multiattribute utility instruments and diseases, as well as estimated using different techniques. This will facilitate preliminary identification of relationships between MID heterogeneity and study characteristics. Meta-regression will be informed using these results.

Meta-regression

A multilevel mixed model, estimated via restricted maximum likelihood estimation, will be used to evaluate sources of MID heterogeneity while controlling for confounding (sources of which are currently indeterminant) and unexplained heterogeneity. Clustering in the data is hypothesised to arise from methods of MID calculation, and the multiattribute utility instruments that MIDs are estimated for. This hypothesis arises from multiattribute utility instruments using different scales and possessing varying levels of sensitivity. The inclusion of the aforementioned levels in the meta-regression model is contingent on hypothesis confirmation. Further details of model specification will be decided after descriptive analysis and subsequent backward elimination of irrelevant variables.

Restricted maximum likelihood estimation is preferred over iterative maximum likelihood approaches which ignore variability in fixed effects and degrees of freedom consumption, during coefficient estimation.[38] Notably, a small sample is expected in the proposed meta-analysis due to the limited number of articles recovered during prestudy, ad hoc database searches. Consequently, disregarded degrees of freedom consumption would likely invalidate statistical inferences in the meta-regression. To maximise the accuracy of statistical inference, restricted maximum likelihood estimation will be paired with the Kenward-Roger small sample correction.[39]

MID pooling

A linear random effects model will be applied to subsets of MIDs, such as those associated with specific multiattribute utility instruments or diseases. This will facilitate the pooling of MID estimates to create multiattribute utility instrument- and methodology-specific legacy MIDs (or legacy anchors). Combined with knowledge of contributors to MID heterogeneity, these legacy MIDs can be used as standards against which MID estimates may be compared.

ETHICS AND DISSEMINATION

Ethics approval is not required for this systematic review, as it intends to analyse existing works. The primary method of study dissemination will be publication in a peer-reviewed journal. Secondary methods of distribution will include presentations at conferences and seminars.

Acknowledgements The authors acknowledge research librarian Mrs Michaela Venn for her assistance in creating the search strategy presented in this protocol.

Contributors This protocol was conceived and initially drafted by GJH and JAC. The associated database search strategy was developed by GJH and JAC. The coauthors (BVT, IvdM, SBC, SS-Y, AJP, QX, BA and AS) reviewed the initial and subsequent drafts, providing substantial suggestions and commentary, with the consequent revisions implemented by GJH. Work undertaken by GJH was performed under the supervision of JAC, and JAC will be the guarantor of the proposed systematic review and meta-analysis. All authors have approved the submission.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs

Glen James Henson http://orcid.org/0000-0002-1237-5710
Bruce V Taylor http://orcid.org/0000-0003-2807-0070
Ingrid van der Mei http://orcid.org/0000-0001-9009-7472
Suzi B Claflin http://orcid.org/0000-0001-6545-946X
Steve Simpson-Yap http://orcid.org/0000-0001-6521-3056
Andrew J Palmer http://orcid.org/0000-0002-9703-7891
Qing Xia http://orcid.org/0000-0002-8556-3456
Benny Antony http://orcid.org/0000-0001-8704-6084

Ambrish Singh http://orcid.org/0000-0002-4618-7507
Julie A Campbell http://orcid.org/0000-0002-1820-6758

REFERENCES

1 Campbell JA, Palmer AJ, Venn A, *et al*. A head-to-head comparison of the EQ-5D-5L and AQoL-8D Multi-Attribute utility instruments in patients who have previously undergone bariatric surgery. *Patient* 2016;9:311–22.
2 Kennedy-Martin M, Slaap B, Herdman M, *et al*. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTa) guidelines. *Eur J Health Econ* 2020;21:1245–57.
3 Richardson J, Sinha K, Iezzi A, *et al*. Modelling utility weights for the Assessment of Quality of Life (AQoL)-8D. *Qual Life Res* 2014;23:2395–404.
4 Herdman M, Gudex C, Lloyd A, *et al*. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;20:1727–36.
5 Kaplan RM, Anderson JP, Ganiats TG. The Quality of Well-being Scale: rationale for a single quality of life index. In: Walker SR, Rosser RM, eds. *Quality of life assessment: key issues in the 1990s*. Dordrecht: Springer Netherlands, 1993: 65–94.
6 Brazier JE, Mulhern BJ, Bjorner JB, *et al*. Developing a new version of the SF-6D health state classification system from the SF-36v2: SF-6Dv2. *Med Care* 2020;58:557–65.
7 Furlong WJ, Feeny DH, Torrance GW, *et al*. The health Utilities index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med* 2001;33:375–84.
8 Drummond MF, Sculpher MJ, Claxton K, *et al*. *Methods for the economic evaluation of health care programmes*. 4th ed. Oxford (UK): Oxford university press, 2015: 123–80.
9 Jin X, Liu GG, Gerstein HC, *et al*. Minimally important difference and predictors of change in quality of life in type 2 diabetes: a community-based survey in China. *Diabetes Metab Res Rev* 2018;34:e3053.
10 Ousmen A, Touraine C, Deliu N, *et al*. Distribution- and anchor-based methods to determine the minimally important difference on patient-reported outcome questionnaires in oncology: a structured review. *Health Qual Life Outcomes* 2018;16:228.
11 Luo N, Johnson J, Coons SJ. Using instrument-defined health state transitions to estimate minimally important differences for four preference-based health-related quality of life instruments. *Med Care* 2010;48:365–71.
12 Zervos TM, Asmaro K, Air EL. Contemporary analysis of minimal clinically important difference in the neurosurgical literature. *Neurosurgery* 2021;88:713–9.
13 Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005;14:1523–32.
14 Wang Y-C, Hart DL, Stratford PW, *et al*. Baseline dependency of minimal clinically important improvement. *Phys Ther* 2011;91:675–88.
15 Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. New York (US): Lawrence Erlbaum Associates, 1988: 19–22.
16 Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;27:S178–89.
17 Johnston BC, Thorlund K, Schünemann HJ, *et al*. Improving the interpretation of quality of life evidence in meta-analyses: the application of minimal important difference units. *Health Qual Life Outcomes* 2010;8:116.
18 Lee AC, Driban JB, Price LL, *et al*. Responsiveness and Minimally Important Differences for 4 Patient-Reported Outcomes Measurement Information System Short Forms: Physical Function, Pain Interference, Depression, and Anxiety in Knee Osteoarthritis. *J Pain* 2017;18:1096–110.
19 Yost KJ, Sorensen MV, Hahn EA, *et al*. Using multiple anchor- and distribution-based estimates to evaluate clinically meaningful change on the functional assessment of cancer Therapy-Biologic response modifiers (FACT-BRM) instrument. *Value Health* 2005;8:117–27.
20 Jayadevappa R, Cook R, Chhatre S. Minimal important difference to infer changes in health-related quality of life-a systematic review. *J Clin Epidemiol* 2017;89:188–98.
21 Devji T, Carrasco-Labra A, Guyatt G. Mind the methods of determining minimal important differences: three critical issues to consider. *Evid Based Ment Health* 2021;24:77–81.
22 Mouelhi Y, Jouve E, Castelli C, *et al*. How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. *Health Qual Life Outcomes* 2020;18:136.
23 Carrasco-Labra A, Devji T, Qasim A, *et al*. Minimal important difference estimates for patient-reported outcomes: a systematic survey. *J Clin Epidemiol* 2021;133:61–71.
24 Bohannon RW, Crouch R. Minimal clinically important difference for change in 6-minute walk test distance of adults with pathology: a systematic review. *J Eval Clin Pract* 2017;23:377–81.
25 Concoff A, Rosen J, Fu F, *et al*. A comparison of treatment effects for nonsurgical therapies and the minimum clinically important difference in knee osteoarthritis: a systematic review. *JBJS Rev* 2019;7:e5.
26 Moher D, Shamseer L, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
27 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
28 Husereau D, Drummond M, Augustovski F, *et al*. Consolidated health economic evaluation reporting standards (cheers) 2022 explanation and elaboration: a report of the ISPOR cheers II good practices Task force. *Value Health* 2022;25:10–31.
29 Wolowacz SE, Briggs A, Belozeroff V, *et al*. Estimating health-state utility for economic models in clinical studies: an ISPOR good research practices Task force report. *Value Health* 2016;19:704–19.
30 Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
31 Terwee CB, Mokkink LB, Knol DL, *et al*. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651–7.
32 Mokkink LB, de Vet HCW, Prinsen CAC, *et al*. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1171–9.
33 Prinsen CAC, Mokkink LB, Bouter LM, *et al*. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.
34 Whiting P, Savović J, Higgins JPT, *et al*. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016;69:225–34.
35 Guyatt GH, Oxman AD, Vist GE, *et al*. Grade: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
36 King MT. A point of minimal important difference (mid): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011;11:171–84.
37 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–15.
38 McNeish D. Small sample methods for multilevel modeling: a colloquial elucidation of REML and the Kenward-Roger correction. *Multivariate Behav Res* 2017;52:661–70.
39 Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997;53:983–97.
40 Sintonen H. The 15D instrument of health-related quality of life: properties and applications. *Ann Med* 2001;33:328–36.
41 Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care* 2004;42:851–9.
42 Chen G, Olsen JA. Filling the psycho-social gap in the EQ-5D: the empirical support for four bolt-on dimensions. *Qual Life Res* 2020;29:3119–29.
43 Dewitt B, Feeny D, Fischhoff B, *et al*. Estimation of a Preference-Based Summary Score for the Patient-Reported Outcomes Measurement Information System: The PROMIS®-Preference (PROPr) Scoring System. *Med Decis Making* 2018;38:683–98.