BMJ Open

# A protocol for a systematic review and meta-analysis of minimum important differences for generic multi-attribute utility instruments



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2022-062703 |
| Article Type: | Protocol |
| Date Submitted by the Author: | 21-Mar-2022 |
| Complete List of Authors: | Henson, Glen; University of Tasmania Menzies Institute for Medical Research, Health Economics<br>Taylor, Bruce; University of Tasmania Menzies Institute for Medical Research<br>van der Mei, Ingrid; University of Tasmania Menzies Institute for Medical Research<br>Claflin, Suzi; University of Tasmania Menzies Institute for Medical Research<br>Simpson-Yap, Steve ; The University of Melbourne School of Population and Global Health,<br>Palmer, Andrew; University of Tasmania Menzies Institute for Medical Research<br>Xia, Qing; University of Tasmania, Menzies Institute for Medical Research<br>Antony, Benny; University of Tasmania, Menzies Institute for Medical Research<br>Singh, Ambrish; University of Tasmania Menzies Institute for Medical Research<br>Campbell, Julie; University of Tasmania Menzies Institute for Medical Research |
| Keywords: | HEALTH ECONOMICS, STATISTICS & RESEARCH METHODS, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT |
| | |

# A protocol for a systematic review and meta-analysis of minimum important differences for generic multi-attribute utility instruments

*Corresponding Author:*

Julie A Campbell, Health Economics, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia.
Postal Address: 17 Liverpool St, Hobart, Tasmania, Australia, 7000.
Email: julie.campbell@utas.edu.au

*Authors:*

Glen J Henson, Health Economics, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: glen.henson@utas.edu.au.
Bruce V Taylor, Neurology and Epidemiology, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: bruce.taylor@utas.edu.au
Ingrid van der Mei, Epidemiology, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: ingrid.vandermei@utas.edu.au
Suzi Claflin, Epidemiology, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: suzi.claflin@utas.edu.au
Steve Simpson-Yap, Epidemiology and Neuroepidemiology, Melbourne School of Population and Global Health (University of Melbourne), Melbourne, Australia. steve.simpsonyap@unimelb.edu.au
Andrew J Palmer, Health Economics, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: andrew.palmer@utas.edu.au
Qing Xia, Health Economics, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. qing.xia@utas.edu.au
Benny Eathakkattu Antony, Musculoskeletal and Epidemiology, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. benny.eathakkattuantony@utas.edu.au
Ambrish Singh, Musculoskeletal and Epidemiology, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: ambrish.singh@utas.edu.au
Julie A Campbell, Health Economics, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: julie.campbell@utas.edu.au

*2646 Words*

# ARTICLE SUMMARY


## Abstract

*Introduction:* Generic multi-attribute utility instruments (MAUIs) are efficient tools for determining and enumerating health-related quality of life. MAUIs accomplish this by mapping patient-reported responses – to arrays of survey questions – to health state utilities (HSUs) via algorithms. Minimum 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 to 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 ten databases, and will be performed and reported in accordance with several validated guidelines, principally the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. 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 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 pertaining to MID heterogeneity. Together with extracted data, this will allow for MID heterogeneity, and its sources, 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. Pooled MIDs may be used in benchmarking new MID estimates.

*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.


*Keywords: Health Economics, Statistics and Research Methods, Protocols and Guidelines*

## Strengths and limitations

- The proposed systematic review and meta-analysis will assist in the establishment of guidelines for the estimation and use of minimum important differences (MIDs). Currently no such guidelines, validated through systematic review and meta-analysis, exist.
- The pending systematic review will investigate ten databases (both biomedical and economic) and apply a broad range of search terms. This will minimise the risk of study omission.
- If successful, the establishment of validated guidelines for MIDs will inform effective and uniform usage of MIDs. This will allow greater certainty, transparency, and comparability in related literature.
- Studies meeting the systematic review's inclusion criteria may be missed, despite a comprehensive search strategy.

# 1.0 INTRODUCTION

The following is a protocol for a systematic review and meta-analysis of minimum important differences (MIDs) for generic multi-attribute utility instruments (MAUIs). This protocol is registered with PROSPERO (number CRD4202126182).

## 1.1 Multi-Attribute utility instruments

Multi-Attribute utility instruments (MAUIs) operate by applying algorithms to arrays of patient-reported outcomes, associated with MAUI survey questions [1]. These algorithms generate health state utilities (HSUs), which are ordinal rankings of health-related quality of life (HRQoL) [2]. To formulate the algorithms, vectors of question responses are mapped to HSUs using a variety of experimental economics techniques including standard gambles (SGs), visual-analogue scales (VASs) and time trade-offs (TTOs) [3]. HSUs are applied in cost-utility analyses (a type of comprehensive health economic analysis, used to evaluate medical interventions), clinical assessments, and evaluation of patient-reported outcomes [1 4].

MAUI surveys pose questions pertaining to 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) [5] MAUI coalesces dimensional scores into super-dimensional scores, which provide a measures of overall physical and mental health. Other common MAUIs include the European Quality of Life – 5 Dimensions – 5 Levels (EQ-5D-5L) [6], Quality of Wellbeing (QWB) [7], Short Form – 6 Dimensions Version 1 (SF-6Dv1) [8], and Health Utilities Index Version 3 (HUI3) [9], which all vary in size and the health dimensions they assess. See *Table 1* for a list of common MAUIs, the dimensions of health they analyse, and the number of items (questions) in each.

Table 1: Health dimensions assessed by eight multi-attribute utility instruments, and the number of items in each

| *Instrument Name* | *Health Dimensions Assessed* | *Number of Items* |
|---|---|---|
| *EQ-5D-5L* [6] | • Mobility<br>• Self-Care<br>• Usual Activities<br>• Pain/Discomfort<br>• Anxiety/ Depression | 5 |
| *AQoL-8D* [5] | • Independent Living<br>• Senses<br>• Pain<br>• Mental Health<br>• Happiness<br>• Self-Worth<br>• Coping<br>• Relationships | 35 |
| *HUI3* [9] | • Vision<br>• Hearing<br>• Speech<br>• Ambulation<br>• Dexterity<br>• Emotion<br>• Cognition<br>• Pain | 8 |
| *QWB* [7] | • Chronic Symptoms<br>• Acute Symptoms<br>• Mental Health<br>• Mobility<br>• Usual Activity<br>• Physical Activity | 74 |
| *15-D* [10] | • Breathing<br>• Mental function<br>• Speech (Communication)<br>• Vision<br>• Mobility<br>• Usual activities<br>• Vitality<br>• Hearing<br>• Eating<br>• Elimination<br>• Sleeping<br>• Distress<br>• Discomfort and Symptoms<br>• Sexual Activity<br>• Depression | 15 |
| *SF-6Dv1* [11] | • Physical Function<br>• Role Limitation<br>• Social Function<br>• Bodily Pain<br>• Mental Health<br>• Vitality | 6 |
| *EQ-5D-5L Psychosocial/ H9-D* [12] | • Mobility<br>• Self-Care<br>• Usual Activities<br>• Pain/Discomfort<br>• Anxiety/ Depression<br>• Vitality<br>• Sleep<br>• Social Relationships<br>• Community Connectedness | 9 |
| *PROPr Scoring System for the PROMIS* [13] | • Cognitive Function<br>• Depression<br>• Fatigue<br>• Pain Interference<br>• Physical Function<br>• Sleep Disturbance<br>• Social Roles and Activities | Variable |

Abbreviations (not appearing previously): 15 – Dimension (15-D), Health – 9 Dimensions (H-9D), PROMIS Preference (PROPr), Patient-Reported Outcome Measurement Information System (PROMIS).

### 1.2 Minimum important differences

Although variations in HRQoL can be measured using MAUIs, these instruments provide no evaluation of what constitutes a clinically significant/meaningful change. Therefore, MIDs – sometimes referred to as minimum *clinically* important differences – are required [14]. These values are the smallest change in HSU that is statistically significant and represents a meaningful adjustment to patient HRQoL [15]. MIDs can lack robustness across MAUIs and populations [16-18].

### 1.3 MID calculation methods

Major methods of MID estimation are described as distribution-based and anchor-based [14]. Distribution-based methods rely on statistical techniques to develop MIDs. An example of such a method is Cohen's effect sizes [19]. Cohen's effect sizes are calculated as $ES = (M_2 - M_1)/S_1$ [19]. In this equation, $M_1$ is the average baseline HSU for a sample of participants. $M_2$ is an HSU greater than the average baseline HSU, which represents, comparatively, a superior health state. $S_1$ is the standard deviation for the mean, baseline HSU. Using a classification scale, the output of the equation can be used to classify a change in HSU as large (not an MID) or small (possibly a MID) [20]. Other distribution-based methods include using fractions of the standard error of the mean as MIDs [3].

Anchor-based methods can be subdivided into external and internal anchors. External anchors involve respondents being separately questioned, following MAUI implementation, regarding whether changes in their HSU represent meaningful changes in their health [14-21]. Contrastingly, internal anchors are instrument-defined. They are derived as the difference in attributable HSUs between two minimally different health states, which are thought to be clinically distinct [15].

Other methods of MID calculation include using legacy anchors, triangulation, and the Delphi method. Legacy anchors are MIDs sourced from previous work and either reapplied to a new study or used to benchmark new MIDs [22]. Triangulation involves use of both distribution and anchor-based methods to generate a single MID [23]. MID triangulation is intended to provide increased internal validity to MID estimates [23]. Lastly, the Delphi method involves establishing MIDs by consensus.

### 1.4 Gaps in the literature

No study has been conducted which is a specific and systematic review and meta-analysis of MIDs for generic MAUIs. Due to this evidence gap, there are also no guidelines regarding MID estimation for generic MAUIs which are validated by a systematic review and meta-analysis. Existing literature has either reviewed MIDs for MAUIs in conjunction with MIDs for disease or symptom-specific instruments [24-27], or focused on MIDs relevant to a particular intervention or disorder [28 29]. Studies applicable to the former category have often been limited in scope, searching few databases [24 26]. Others such studies had different aims to guideline construction, such as highlighting research gaps through systematic review [25], or establishing an MID repository [27].

### 1.5 Research questions

The proposed systematic review and meta-analysis will address the following research questions regarding MIDs for MAUIs:

1. How were MIDs calculated:
    a. Which distribution or anchor-based methods were applied?
    b. Which methods are most commonly used?
    c. Were the methods novel and if so in what way?
2. For what MAUIs and diseases were MIDs calculated:
    a. Were MIDs consistent across MAUIs and diseases?
    b. Is variance present in MIDs across iterations using the same, similar, or different cohorts?
    c. Can existing MIDs be applied to new research, and under what circumstances?
3. Are methods of MID estimation theoretically and empirically sound:
    a. Were there any mathematical errors or controversial innovations?
    b. How, if at all, were the methods validated?
    c. Did different calculation methods produce significantly different MIDs?
4. How were MIDs evaluated:
    a. What, if any, guidelines were used to evaluate MIDs and were these guidelines validated?
    b. What was the result of MID validations?
5. What variables, if any, contribute systematically to heterogeneity in MID estimates:
    a. Can regression-based evidence be acquired to support relevant associations?

b. If influential variables are controlled for, do MID estimates converge?
c. What level of unexplained heterogeneity exists?

### 1.6 Aim and rationale

The aim of the review is to generate complete and nuanced guidelines to MIDs for generic MAUIs, 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.

## 2.0 METHODS AND ANALYSIS

### 2.1 Patient and public involvement

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

### 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 International Society for Pharmacoeconomics Outcomes Research (ISPOR) good research practices taskforce report regarding HSUs in clinical studies [32].

### 2.3 Validated guidelines: quality appraisal

The Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist will be adopted to determine the suitability of studies, meeting inclusion criteria, for incorporation into the systematic review [33]. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) patient-reported outcome measures assessment methodology will be applied to evaluate the quality of included papers [34]. To evaluate the quality of the systematic review and meta-analysis, the COSMIN guidelines for systematic reviews of patient-reported outcomes were chosen [35]. Additionally, references from included papers will be screened for relevant articles to identify potential omissions in the systematic review.

### 2.4 Validated guidelines: risk of bias assessment

The COSMIN risk of bias checklist will be used to evaluate potential bias in studies meeting the inclusion criteria [36]. Any studies found to be at high risk of bias will be weighted in meta-analysis, to reduce their impact on review conclusions. To assess risk of bias in the systematic review and meta-analysis, the Risk of BIas assessment tool for Systematic reviews (ROBIS) was selected [37]. 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) [37].

### 2.5 Search methodology

A pre-study, 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 1*.

The search strategy selected requires one word or phrase from each of the 'minimal important difference' divisions and phrase or name from either 'multiple attribute utility instrument' division to be present in an article's title and/or abstract for that paper to be considered for inclusion. Additionally, search terms will be trialled as pluralised (hyphenated) and singular (non-hyphenated) variants. Relevant acronyms are to be applied in searches, as well as their respective expansions.

Both economic and biomedical electronic databases will be searched in this review, from 01/04/2022. Economic databases to be investigated are the American Economic Association (EconLit), Ideas, the Centre for Reviews and Dissemination (CRD), which includes the Database of Abstracts of Reviews of Effects (DARE) and National Health Service Economic Evaluation Database (NHS EED), the Health Technology Assessment Database (HTA), and the Cost-Effectiveness Analysis Registry (CEA Registry). Biomedical databases that are to be examined include PubMed, PsycINFO, CINHAL, Patient-Reported Outcome and Quality of Life Instruments Database, and Embase via Ovid. In addition, google scholar will be utilised to maximise the completeness of the review.

Figure 1 – Synonymic groupings of search terms

## 2.6 Inclusion criteria

This review will include English papers which incorporate MIDs for generic MAUIs that generate HSUs. 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 [38]. Furthermore, only original, published studies will be included; editorials, commentaries, protocols, reviews, unpublished works, and meta-analyses are to be excluded. In vitro and animal studies will also be excluded.

## 2.7 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 screen accumulated papers through analysis of titles and abstracts, excluding those not meeting the inclusion criteria (detailed in section 2.1). A 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, co-authors will be invited to mediate.

## 2.8 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 is not present in a paper, authors will be contacted. The following data will be extracted from included studies:

1. Clinical and sociodemographic statistics for study samples: age, income, sex, education, urbanity of residence, insurance, number of subjects, diseases and comorbidities, national health service, country, response rates, medication use, disease phenotype, and representativeness of samples.

2. Instruments used: which instruments, instrument versions (such as HUI2 and HUI3) and variations (for example, the EQ-5D-3L and EQ-5D-5L) were applied.

3. Publication characteristics: first author, date, journal, country of origin, study design, quality, and adherence to validated guidelines.

4. Mathematics: methods of MID calculation (distribution-based, anchor-based, or novel) and validation methods (such as triangulation through multiple methods and statistical inference).

5. Sample selection: representativeness assurance method, exclusion and inclusion criteria, subject retention, and matching strategies.

6. Results: MIDs, standard errors, robustness, stated comparisons to previous work (for example, were results considered inferior or superior and why), and discussions regarding strengths and limitations.

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

### 2.10 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.

### 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. 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 MAUIs and diseases, as well as estimated using different techniques. This will facilitate preliminary identification of relationships between MID heterogeneity and study characteristics. Elements of meta-regression will be informed using these results.

### 2.12 Meta-regression

A multilevel mixed model, estimated via restricted maximum likelihood estimation (REML), will be used to evaluate sources of MID heterogeneity whilst 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 MAUIs that MIDs are estimated for. This hypothesis arises from MAUIs using different scales and possessing varying levels of sensitivity. 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.

REML estimation is preferred over iterative maximum likelihood approaches which ignore variability in fixed effects and degrees of freedom consumption, during coefficient estimation [39]. Notably, a small sample is expected in the proposed meta-analysis due to the limited number of articles recovered during pre-study, ad-hoc database searches. Consequently, disregarded degrees of freedom consumption would likely invalidate statistical inferences pertaining to the meta-regression. To maximise the accuracy of statistical inference, REML estimation will be paired with the Kenward-Roger small sample correction [40].

### 2.13 MID Pooling

A linear random effects model will be applied to subsets of MIDs, such as those associated with specific MAUIs or diseases. This will facilitate the pooling of MID estimates to create MAUI 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.

## 3.0 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 presentation at conferences and seminars.

# 4.0 ADDITIONAL

## 4.1 Authors' Contributions

This protocol was conceived of and initially drafted by GJH and JAC. The associated database search strategy was developed by GJH and JAC in consultation with librarian Michaela Venn. The co-authors (BVT, IM, SC, SS, AJP, QX, BEA, 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 submission.

## 4.2 Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

## 4.3 Conflicts of interests

All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/disclosure-of-interest/ and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

## 4.4 Data sharing statement

Data resulting from the proposed systematic review will be published with the review.

## 5.0 REFERENCES

1. Campbell JA, Palmer AJ, Venn A, Sharman M, Otahal P, Neil A. 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(4):311-22.doi:10.1007/s40271-015-0157-5

2. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. 4th ed. Oxford (UK): Oxford university press 2015:123-180.

3. Jin X, Liu GG, Gerstein HC, Levine MAH, Guan H, Li H, 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(8):e3053.doi:10.1002/dmrr.3053

4. Kennedy-Martin M, Slaap B, Herdman M, van Reenen M, Kennedy-Martin T, Greiner W, 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(8):1245-57.doi:10.1007/s10198-020-01195-8

5. Richardson J, Sinha K, Iezzi A, Khan MA. Modelling utility weights for the Assessment of Quality of Life (AQoL)-8D. *Qual Life Res*. 2014;23(8):2395-404.doi:10.1007/s11136-014-0686-8

6. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-36.doi:10.1007/s11136-011-9903-x

7. 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.

8. Brazier JE, Mulhern BJ, Bjorner JB, Gandek B, Rowen D, Alonso J, et al. Developing a new version of the SF-6D health state classification system from the SF-36v2: SF-6Dv2. *Med Care*. 2020;58(6):557-65.doi:10.1097/MLR.0000000000001325

9. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med*. 2001;33(5):375-84.doi:10.3109/07853890109002092

10. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. *Ann Med*. 2001;33(5):328-36.doi:10.3109/07853890109002086

11. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care*. 2004;42(9):851-9.doi:10.1097/01.mlr.0000135827.18610.0d

12. 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(11):3119-29.doi:10.1007/s11136-020-02576-5

13. Dewitt B, Feeny D, Fischhoff B, Cella D, Hays RD, Hess R, 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(6):683-98.doi:10.1177/0272989X18776637

14. Ousmen A, Touraine C, Deliu N, Cottone F, Bonnetain F, Efficace F, 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(1):228.doi:10.1186/s12955-018-1055-z

15. 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(4):365-71.doi:10.1093/neuros/nyaa490

16. Zervos TM, Asmaro K, Air EL. Contemporary analysis of minimal clinically important difference in the neurosurgical literature. *Neurosurgery*. 2021;88(4):713-9.doi:10.1007/s11136-004-7713-0

17. 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(6):1523-32.doi:10.1007/s11136-004-7713-0

18. Wang YC, Hart DL, Stratford PW, Mioduski JE. Baseline dependency of minimal clinically important improvement. *Phys Ther*. 2011;91(5):675-88.doi:10.2522/ptj.20100229

19. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. New York (US): Lawrence Erlbaum Associates 1988:19-22.

20. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care*. 1989;27(3 Suppl):S178-89.doi:10.1097/00005650-198903001-00015

21. Johnston BC, Thorlund K, Schunemann HJ, Xie F, Murad MH, Montori VM, 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(1):116.doi:10.1186/1477-7525-8-116

22. Lee AC, Driban JB, Price LL, Harvey WF, Rodday AM, Wang C. 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(9):1096-110.doi:10.1016/j.jpain.2017.05.001

23.Yost KJ, Sorensen MV, Hahn EA, Glendenning GA, Gnanasakthy A, Cella D. 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(2):117-27.doi:10.1111/j.1524-4733.2005.08202.x

24. 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.doi:10.1016/j.jclinepi.2017.06.009

25. 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(2):77-81.doi:10.1136/ebmental-2020-300164

26. Mouelhi Y, Jouve E, Castelli C, Gentile S. 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;18:1-17.doi.10.1186/s12955-020-01344-w

27. Carrasco-Labra A, Devji T, Qasim A, Phillips MR, Wang Y, Johnston BC, et al. Minimal important difference estimates for patient-reported outcomes: a systematic survey. *J Clin Epidemiol*. 2021;133:61-71.doi:10.1016/j.jclinepi.2020.11.024

28. 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(2):377-81.doi:10.1111/jep.12629

29. Concoff A, Rosen J, Fu F, Bhandari M, Boyer K, Karlsson J, 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(8):e5.doi:10.2106/JBJS.RVW.18.00150

30. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4(1):1.doi:10.1186/2046-4053-4-1

31. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.doi:10.1371/journal.pmed.1000097

32. Wolowacz SE, Briggs A, Belozeroff V, Clarke P, Doward L, Goeree R, et al. Estimating health-state utility for economic models in clinical studies: an ISPOR Good Research Practices Task Force report. *Value Health*. 2016;19(6):704-19.doi:10.1016/j.jval.2016.06.001.

33. Husereau D, Drummond M, Augustovski F, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) explanation and elaboration: a report of the ISPOR CHEERS II Good Practices Task Force. *Value Health*. 2022;25(1):10-31.doi:10.1016/j.jval.2021.10.008

34. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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(4):539-49.

35. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-57.doi:10.1007/s11136-018-1798-3

36. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171-9.doi:10.1007/s11136-017-1765-4

37. Whiting P, Savovic J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2016;69:225-34.doi:10.1016/j.jclinepi.2015.06.005

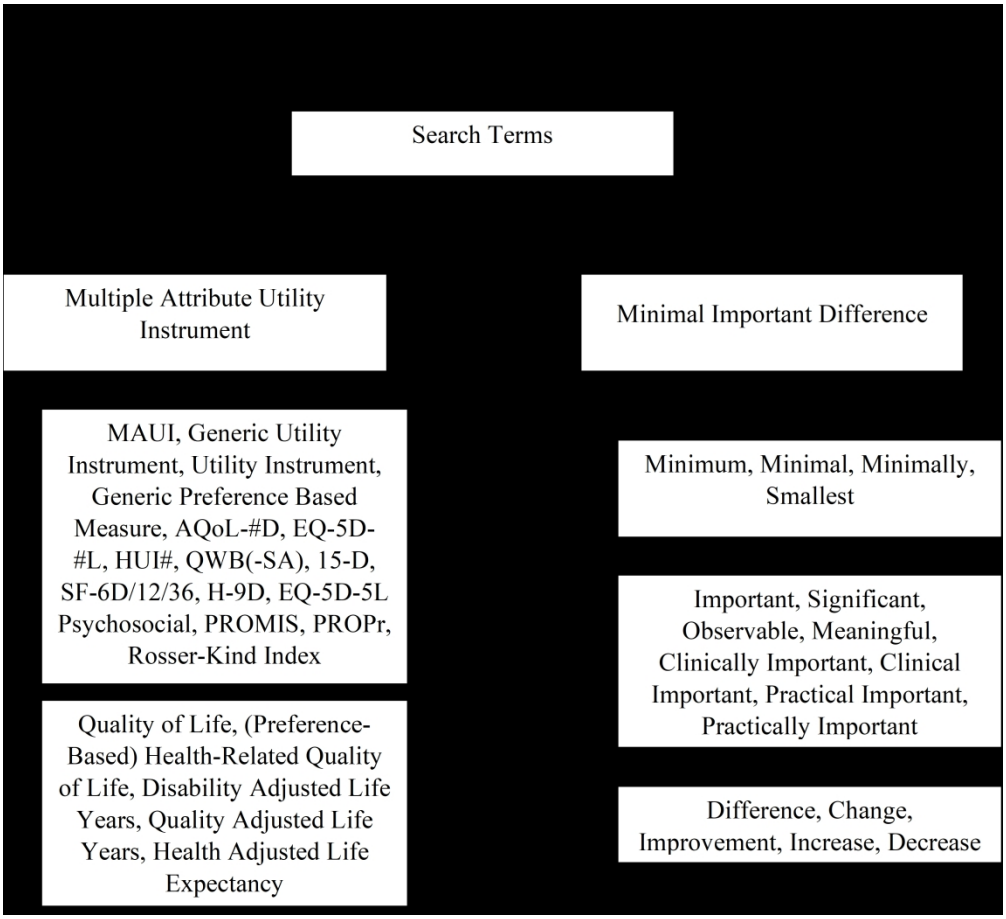
38. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-15.doi:10.1016/0197-2456(89)90005-6

39. McNeish D. Small sample methods for multilevel modeling: a colloquial elucidation of REML and the Kenward-Roger correction. *Multivariate Behav Res*. 2017;52(5):661-70.doi:10.1080/00273171.2017.1344538

40. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*. 1997;53(3):983-97.doi:https://doi.org/10.2307/2533558

Figure 1 Caption

Words associated with 'minimum important difference' are divided by which element of the phrase they are synonymous with. From top to bottom, synonyms are associated with minimum, important, and difference. Furthermore, where short phrases are included as synonyms, the second element of these phrases are to be interchanged with individual synonyms. Words associated with 'multiple attribute utility instrument' are divided into instrument names and outcome measures associated with MAUIs.




Words associated with 'minimum important difference' are divided by which element of the phrase they are synonymous with. From top to bottom, synonyms are associated with minimum, important, and difference. Furthermore, where short phrases are included as synonyms, the second element of these phrases are to be interchanged with individual synonyms. Words associated with 'multiple attribute utility instrument' are divided into instrument names and outcome measures associated with MAUIs.

404x368mm (236 x 236 DPI)

# PRISMA-P 2015 Checklist

| Section/topic | # | Checklist item | Information reported | | Line number(s) |
|---|---|---|---|---|---|
| | | | Yes | No | |
| **ADMINISTRATIVE INFORMATION** | | | | | |
| **Title** | | | | | |
| Identification | 1a | Identify the report as a protocol of a systematic review | X | | Page 4, lines 1 |
| Update | 1b | If the protocol is for an update of a previous systematic review, identify as such | NA | | |
| **Registration** | 2 | If registered, provide the name of the registry (e.g., PROSPERO) and registration number in the Abstract | X | | Page 4, lines 2-3 |
| **Authors** | | | | | |
| Contact | 3a | Provide name, institutional affiliation, and e-mail address of all protocol authors; provide physical mailing address of corresponding author | X | | Page 1, lines 3-28 |
| Contributions | 3b | Describe contributions of protocol authors and identify the guarantor of the review | X | | Page 13, lines 2-9 |
| **Amendments** | 4 | If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments | NA | | |
| **Support** | | | | | |
| Sources | 5a | Indicate sources of financial or other support for the review | X | | Page 13, lines 10-22 |
| Sponsor | 5b | Provide name for the review funder and/or sponsor | NA | | |
| Role of sponsor/funder | 5c | Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol | NA | | |
| **INTRODUCTION** | | | | | |
| **Rationale** | 6 | Describe the rationale for the review in the context of what is already known | X | | Pages 7(8), lines 1-10(3-10) |
| **Objectives** | 7 | Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO) | X | | Pages 7(8), lines 11-32(1-2) |
| **METHODS** | | | | | |
| **Eligibility criteria** | 8 | Specify the study characteristics (e.g., PICO, study design, setting, time frame) and report characteristics (e.g., years considered, language, publication status) to be used as criteria for eligibility for the review | X | | Page 10, lines 2-9 |
| **Information sources** | 9 | Describe all intended information sources (e.g., electronic databases, contact with study authors, trial registers, or other grey literature sources) with planned dates of coverage | X | | Pages 9, lines 23-31 |
| **Search strategy** | 10 | Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated | X | | Page 9, lines 13-22 |

| Section/topic | # | Checklist item | Information reported | | Line number(s) |
|---|---|---|---|---|---|
| | | | Yes | No | |
| ***STUDY RECORDS*** | | | | | |
| Data management | 11a | Describe the mechanism(s) that will be used to manage records and data throughout the review | X | | Page 12, lines 9-13 |
| Selection process | 11b | State the process that will be used for selecting studies (e.g., two independent reviewers) through each phase of the review (i.e., screening, eligibility, and inclusion in meta-analysis) | X | | Page 11, lines 9-17 |
| Data collection process | 11c | Describe planned method of extracting data from reports (e.g., piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators | X | | Page 10, lines 19-22 |
| **Data items** | 12 | List and define all variables for which data will be sought (e.g., PICO items, funding sources), any pre-planned data assumptions and simplifications | X | | Pages 11(12), lines 23-30(1-8) |
| **Outcomes and prioritization** | 13 | List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale | X | | Page 8, lines 3-10 |
| **Risk of bias in individual studies** | 14 | Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis | X | | Page 9, lines 2-4 |
| ***DATA*** | | | | | |
| **Synthesis** | 15a | Describe criteria under which study data will be quantitatively synthesized | X | | Page 11, lines 21-23 |
| | 15b | If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data, and methods of combining data from studies, including any planned exploration of consistency (e.g., $I^2$, Kendall's tau) | X | | Page 11, lines 14-24 |
| | 15c | Describe any proposed additional analyses (e.g., sensitivity or subgroup analyses, meta-regression) | X | | Pages 11(12), lines 23-29(1-22) |
| | 15d | If quantitative synthesis is not appropriate, describe the type of summary planned | NA | | |
| **Meta-bias(es)** | 16 | Specify any planned assessment of meta-bias(es) (e.g., publication bias across studies, selective reporting within studies) | X | | Page 9, lines 4-11 |
| **Confidence in cumulative evidence** | 17 | Describe how the strength of the body of evidence will be assessed (e.g., GRADE) | X | | Page 8, lines 20-26 |

BMJ Open

# A protocol for a systematic review and meta-analysis of minimal important differences for generic multi-attribute utility instruments



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2022-062703.R1 |
| Article Type: | Protocol |
| Date Submitted by the Author: | 26-Aug-2022 |
| Complete List of Authors: | Henson, Glen; University of Tasmania Menzies Institute for Medical Research, Health Economics<br>Taylor, Bruce; University of Tasmania Menzies Institute for Medical Research<br>van der Mei, Ingrid; University of Tasmania Menzies Institute for Medical Research<br>Claflin, Suzi; University of Tasmania Menzies Institute for Medical Research<br>Simpson-Yap, Steve ; The University of Melbourne School of Population and Global Health,<br>Palmer, Andrew; University of Tasmania Menzies Institute for Medical Research<br>Xia, Qing; University of Tasmania, Menzies Institute for Medical Research<br>Antony, Benny; University of Tasmania, Menzies Institute for Medical Research<br>Singh, Ambrish; University of Tasmania Menzies Institute for Medical Research<br>Campbell, Julie; University of Tasmania Menzies Institute for Medical Research |
| Primary Subject Heading: | Health economics |
| Secondary Subject Heading: | Health services research |
| Keywords: | HEALTH ECONOMICS, STATISTICS & RESEARCH METHODS, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT |
| | |

# A protocol for a systematic review and meta-analysis of minimal important differences for generic multi-attribute utility instruments

*Corresponding Author:*

Julie A Campbell, Health Economics, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia.
Postal Address: 17 Liverpool St, Hobart, Tasmania, Australia, 7000.
Email: julie.campbell@utas.edu.au

*Authors:*

Glen J Henson, Health Economics, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: glen.henson@utas.edu.au.
Bruce V Taylor, Neurology and Epidemiology, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: bruce.taylor@utas.edu.au
Ingrid van der Mei, Epidemiology, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: ingrid.vandermei@utas.edu.au
Suzi Claflin, Epidemiology, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: suzi.claflin@utas.edu.au
Steve Simpson-Yap, Epidemiology and Neuroepidemiology, Melbourne School of Population and Global Health (University of Melbourne), Melbourne, Australia. steve.simpsonyap@unimelb.edu.au
Andrew J Palmer, Health Economics, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: andrew.palmer@utas.edu.au
Qing Xia, Health Economics, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. qing.xia@utas.edu.au
Benny Eathakkattu Antony, Musculoskeletal and Epidemiology, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. benny.eathakkattuantony@utas.edu.au
Ambrish Singh, Musculoskeletal and Epidemiology, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: ambrish.singh@utas.edu.au
Julie A Campbell, Health Economics, Menzies Institute for Medical Research (University of Tasmania), Hobart, Australia. Email: julie.campbell@utas.edu.au

*2956 Words*

# ARTICLE SUMMARY


## Abstract

*Introduction:* Generic multi-attribute 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 ten databases (EconLit, IDEAs database, INAHTA database, Medline, PsycINFO, Embase, Emcare, JBIEBP, and CINAHL) from 01/06/2022 through 07/06/2022, 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 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 extracted data, this will allow for MID heterogeneity, and its sources, 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.


*Keywords: Health Economics, Statistics and Research Methods, Protocols and Guidelines*

Strengths and limitations

- 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 was chosen for meta-regression to allow for variability in fixed-effects estimates and degree of freedom consumption.
- Use of REML will permit superior statistical inference compared to 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.

# 1.0 INTRODUCTION

The following is a protocol for a systematic review and meta-analysis of minimal important differences (MIDs) for generic multi-attribute utility instruments. This protocol is registered with PROSPERO (number CRD42021261821).

## 1.1 Multi-attribute utility instruments

Multi-attribute utility instruments can be generic, and adopted for use with any study population or 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 $2.4 \times 10^{23}$ 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 EQ-5D-5L is presented to exemplify the operation of a generic multi-attribute utility instrument.

Table 1 – Health dimensions assessed by eight multi-attribute utility instruments, and the number of items in each

| *Instrument Name* | *Health Dimensions Assessed* | *Number of Items* |
|---|---|---|
| *EQ-5D-5L* [4] | • Mobility<br>• Self-Care<br>• Usual Activities<br>• Pain/Discomfort<br>• Anxiety/ Depression | 5 |
| *AQoL-8D* [3] | • Independent Living<br>• Senses<br>• Pain<br>• Mental Health<br>• Happiness<br>• Self-Worth<br>• Coping<br>• Relationships | 35 |
| *HUI3 (Self-Administered)* [7] | • Vision<br>• Hearing<br>• Speech<br>• Ambulation<br>• Dexterity<br>• Emotion<br>• Cognition<br>• Pain | 15 |
| *QWB* [9] | • Chronic Symptoms<br>• Acute Symptoms<br>• Mental Health<br>• Mobility<br>• Usual Activity<br>• Physical Activity | 74 |
| *15-D* [10] | • Breathing<br>• Mental function<br>• Speech (Communication)<br>• Vision<br>• Mobility<br>• Usual activities<br>• Vitality<br>• Hearing<br>• Eating<br>• Elimination<br>• Sleeping<br>• Distress<br>• Discomfort and Symptoms<br>• Sexual Activity<br>• Depression | 15 |
| *SF-6Dv1* [11] | • Physical Function<br>• Role Limitation<br>• Social Function<br>• Bodily Pain<br>• Mental Health<br>• Vitality | 6 |
| *EQ-5D-5L Psychosocial* [12] | • Mobility<br>• Self-Care<br>• Usual Activities<br>• Pain/Discomfort<br>• Anxiety/ Depression<br>• Vitality<br>• Sleep<br>• Social Relationships<br>• Community Connectedness | 9 |
| *PROPr Scoring System for the PROMIS* [13] | • Cognitive Function<br>• Depression<br>• Fatigue<br>• Pain Interference<br>• Physical Function<br>• Sleep Disturbance<br>• Social Roles and Activities | Variable |

Abbreviations (not appearing previously): 15 – Dimension (15-D), PROMIS Preference (PROPr), Patient-Reported Outcome Measurement Information System (PROMIS).

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

## 1.2 Minimal important differences

Although variations in health-related quality of life can be measured using multi-attribute utility instruments, these instruments provide no evaluation of what constitutes a clinically significant/meaningful change. Therefore, MIDs are required [14]. 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 [15]. MIDs can lack robustness across multi-attribute utility instruments and populations [16-18].

## 1.3 MID calculation methods

Major methods of MID estimation are described as distribution-based and anchor-based [14]. Distribution-based methods rely on statistical techniques to develop MIDs. An example of such a method is Cohen's effect sizes [19]. Cohen's effect sizes are calculated as $ES = (M_2 - M_1)/S_1$ [19]. 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 deviation for 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 a MID) or small (possibly a MID) [20]. Other distribution-based methods include using fractions of the standard error of the mean as MIDs [3].

Anchor-based methods can be subdivided into external and internal anchors. External anchors can involve respondents being separately questioned, following multi-attribute utility instrument implementation, regarding whether changes in their health state utility represent meaningful changes in their health [14-21]. 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 [15].

Other methods of MID calculation include using legacy anchors, triangulation, and the Delphi method. Legacy anchors are MIDs sourced from previous work and either reapplied to a new study or used to benchmark new MIDs [22]. Triangulation involves the use of both distribution and anchor-based methods to generate a single MID [23]. MID triangulation is intended to

provide increased internal validity to MID estimates [23]. Lastly, the Delphi method involves establishing MIDs by consensus.

### 1.4 Gaps in the literature

No study has been conducted which is a specific and systematic review and meta-analysis of MIDs for generic multi-attribute utility instruments. Due to this evidence gap, there are also no guidelines regarding MID estimation for generic multi-attribute utility instruments which are validated by a systematic review and meta-analysis. Existing literature has either reviewed MIDs for multi-attribute utility instruments in conjunction with MIDs for disease or symptom-specific instruments[24-27] or focused on MIDs relevant to a particular intervention or disorder [28 29]. Studies applicable to the former category have often been limited in scope, searching few databases [24 26]. Other such studies had different aims than guideline construction, such as highlighting research gaps through systematic review[25] or establishing a MID repository [27].

### 1.5 Research questions

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

1. How were MIDs calculated?
    a. Which methods were applied?
    b. Which methods are most commonly used?
    c. Were some methods novel and if so in what way?
    d. Did different calculation methods produce significantly different MIDs?
2. For what multi-attribute utility instruments and diseases were MIDs calculated?
    a. Were MIDs consistent across multi-attribute utility instruments and diseases?
    b. Is variation present in MIDs across iterations using the same, similar, or different study cohorts?
3. Are applied methods of MID estimation theoretically and empirically sound?
    a. Were there any mathematical errors or controversial innovations?
    b. Were the methods validated?
4. How were MIDs evaluated?
    a. What, if any, guidelines were used to evaluate MIDs and were these guidelines validated?
    b. What was the result of MID evaluations?
5. What variables, if any, contribute systematically to heterogeneity in MID estimates?

a. Can regression-based evidence be acquired to support relevant associations?
b. If influential variables are controlled for, do MID estimates converge?
c. What level of unexplained heterogeneity exists?

6. Can existing MIDs be applied to new research and under what circumstances?

### 1.6 Aim and rationale

The review aims to generate complete and nuanced guidelines for MIDs for generic multi-attribute 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.

## 2.0 METHODS AND ANALYSIS

### 2.1 Patient and public involvement

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

### 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 reviewed 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.

## 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].

## 2.5 Search methodology

A pre-study, 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 'minimal important difference' divisions and a phrase or name from either 'multi-attribute utility instrument' division to be present in an article's title and/or abstract for that paper to be considered for inclusion. Additionally, 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 (minimal important difference' are present in the search strategy due to the heterogeneity of their usage and the lack of a firmly established nomenclature [40]. See the Appendix for the precise search strategy used in all database searches.

Both economic and biomedical electronic databases will be searched in this review, from 01/06/2022 through 07/06/2022. 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.

Figure 2 – Synonymic groupings of search terms

### 2.6 Inclusion criteria

This review will include English papers that incorporate MIDs for generic multi-attribute 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 [41]. 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.

### 2.7 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 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, co-authors will be invited to mediate.

### 2.8 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 is not present in a paper, authors will be contacted. The following will be extracted from included studies:

1. Participant characteristics: age, socio-economic 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 standard errors, 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.

### 2.9 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 (GH) 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.

### 2.10 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.

### 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 (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 multi-attribute utility

instruments and diseases, as well as estimated using different techniques. This will facilitate preliminary identification of relationships between MID heterogeneity and study characteristics. Elements of meta-regression will be informed using these results.

### 2.12 Meta-regression

A multilevel mixed model, estimated via restricted maximum likelihood estimation, will be used to evaluate sources of MID heterogeneity whilst 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 multi-attribute utility instruments that MIDs are estimated for. This hypothesis arises from multi-attribute 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 [42]. Notably, a small sample is expected in the proposed meta-analysis due to the limited number of articles recovered during pre-study, 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 [43].

### 2.13 MID Pooling

A linear random effects model will be applied to subsets of MIDs, such as those associated with specific multi-attribute utility instruments or diseases. This will facilitate the pooling of MID estimates to create multi-attribute utility instruments 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.

## 3.0 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 published in a peer-reviewed journal. Secondary methods of distribution will include presentations at conferences and seminars.

## 4.0 ADDITIONAL

### 4.1 Authors' Contributions

This protocol was conceived and initially drafted by GJH and JAC. The associated database search strategy was developed by GJH and JAC. The co-authors (BVT, IM, SC, SS, AJP, QX, BEA, 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.

### 4.2 Acknowledgements

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

### 4.3 Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

### 4.4 Conflicts of interests

All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/disclosure-of-interest/ and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

### 4.5 Data sharing statement

Data resulting from the proposed systematic review will be published with the review.

## 5.0 REFERENCES


1. Campbell JA, Palmer AJ, Venn A, Sharman M, Otahal P, Neil A. 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(4):311-22.doi:10.1007/s40271-015-0157-5

2. Kennedy-Martin M, Slaap B, Herdman M, van Reenen M, Kennedy-Martin T, Greiner W, 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(8):1245-57.doi:10.1007/s10198-020-01195-8

3. Richardson J, Sinha K, Iezzi A, Khan MA. Modelling utility weights for the Assessment of Quality of Life (AQoL)-8D. *Qual Life Res*. 2014;23(8):2395-404.doi:10.1007/s11136-014-0686-8

4. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-36.doi:10.1007/s11136-011-9903-x

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, Gandek B, Rowen D, Alonso J, et al. Developing a new version of the SF-6D health state classification system from the SF-36v2: SF-6Dv2. *Med Care*. 2020;58(6):557-65.doi:10.1097/MLR.0000000000001325

7. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med*. 2001;33(5):375-84.doi:10.3109/07853890109002092

8. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. 4th ed. Oxford (UK): Oxford university press 2015:123-180.

9. Jin X, Liu GG, Gerstein HC, Levine MAH, Guan H, Li H, 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(8):e3053.doi:10.1002/dmrr.3053

10. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. *Ann Med*. 2001;33(5):328-36.doi:10.3109/07853890109002086

11. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care*. 2004;42(9):851-9.doi:10.1097/01.mlr.0000135827.18610.0d

12. 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(11):3119-29.doi:10.1007/s11136-020-02576-5

13. Dewitt B, Feeny D, Fischhoff B, Cella D, Hays RD, Hess R, 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(6):683-98.doi:10.1177/0272989X18776637

14. Ousmen A, Touraine C, Deliu N, Cottone F, Bonnetain F, Efficace F, 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(1):228.doi:10.1186/s12955-018-1055-z

15. 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(4):365-71.doi:10.1093/neuros/nyaa490

16. Zervos TM, Asmaro K, Air EL. Contemporary analysis of minimal clinically important difference in the neurosurgical literature. *Neurosurgery*. 2021;88(4):713-9.doi:10.1007/s11136-004-7713-0

17. 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(6):1523-32.doi:10.1007/s11136-004-7713-0

18. Wang YC, Hart DL, Stratford PW, Mioduski JE. Baseline dependency of minimal clinically important improvement. *Phys Ther*. 2011;91(5):675-88.doi:10.2522/ptj.20100229

19. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. New York (US): Lawrence Erlbaum Associates 1988:19-22.

20. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care*. 1989;27(3 Suppl):S178-89.doi:10.1097/00005650-198903001-00015

21. Johnston BC, Thorlund K, Schunemann HJ, Xie F, Murad MH, Montori VM, 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(1):116.doi:10.1186/1477-7525-8-116

22. Lee AC, Driban JB, Price LL, Harvey WF, Rodday AM, Wang C. 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(9):1096-110.doi:10.1016/j.jpain.2017.05.001

23.Yost KJ, Sorensen MV, Hahn EA, Glendenning GA, Gnanasakthy A, Cella D. 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(2):117-27.doi:10.1111/j.1524-4733.2005.08202.x

24. 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.doi:10.1016/j.jclinepi.2017.06.009

25. 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(2):77-81.doi:10.1136/ebmental-2020-300164

26. Mouelhi Y, Jouve E, Castelli C, Gentile S. 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;18:1-17.doi.10.1186/s12955-020-01344-w

27. Carrasco-Labra A, Devji T, Qasim A, Phillips MR, Wang Y, Johnston BC, et al. Minimal important difference estimates for patient-reported outcomes: a systematic survey. *J Clin Epidemiol*. 2021;133:61-71.doi:10.1016/j.jclinepi.2020.11.024

28. 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(2):377-81.doi:10.1111/jep.12629

29. Concoff A, Rosen J, Fu F, Bhandari M, Boyer K, Karlsson J, 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(8):e5.doi:10.2106/JBJS.RVW.18.00150

30. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4(1):1.doi:10.1186/2046-4053-4-1

31. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372.doi:10.1136/bmj.n71

32. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) explanation and elaboration: a report of the ISPOR CHEERS II Good Practices Task Force. *Value Health*. 2022;25(1):10-31.doi:10.1016/j.jval.2021.10.008

33. Wolowacz SE, Briggs A, Belozeroff V, Clarke P, Doward L, Goeree R, et al. Estimating health-state utility for economic models in clinical studies: an ISPOR Good Research Practices Task Force report. *Value Health*. 2016;19(6):704-19.doi:10.1016/j.jval.2016.06.001

34. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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(4):539-49.doi:10.1007/s11136-010-9606-8

35. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012 May;21(4):651-7.doi:10.1007/s11136-011-9960-1

36. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171-9.doi:10.1007/s11136-017-1765-4

37. Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res*. 2018 May;27(5):1147-57.doi:10.1007/s11136-018-1798-3

38. Whiting P, Savovic J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2016;69:225-34.doi:10.1016/j.jclinepi.2015.06.005

39. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clinical research ed)*. 2008;336(7650):924-6.doi:10.1136/bmj.39489.470347.AD.

40. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(2):171-184.doi:10.1586/erp.11.9

41. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-15.doi:10.1016/0197-2456(89)90005-6

42. McNeish D. Small sample methods for multilevel modeling: a colloquial elucidation of REML and the Kenward-Roger correction. *Multivariate Behav Res*. 2017;52(5):661-70.doi:10.1080/00273171.2017.1344538

43. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*. 1997;53(3):983-97.doi:10.2307/2533558

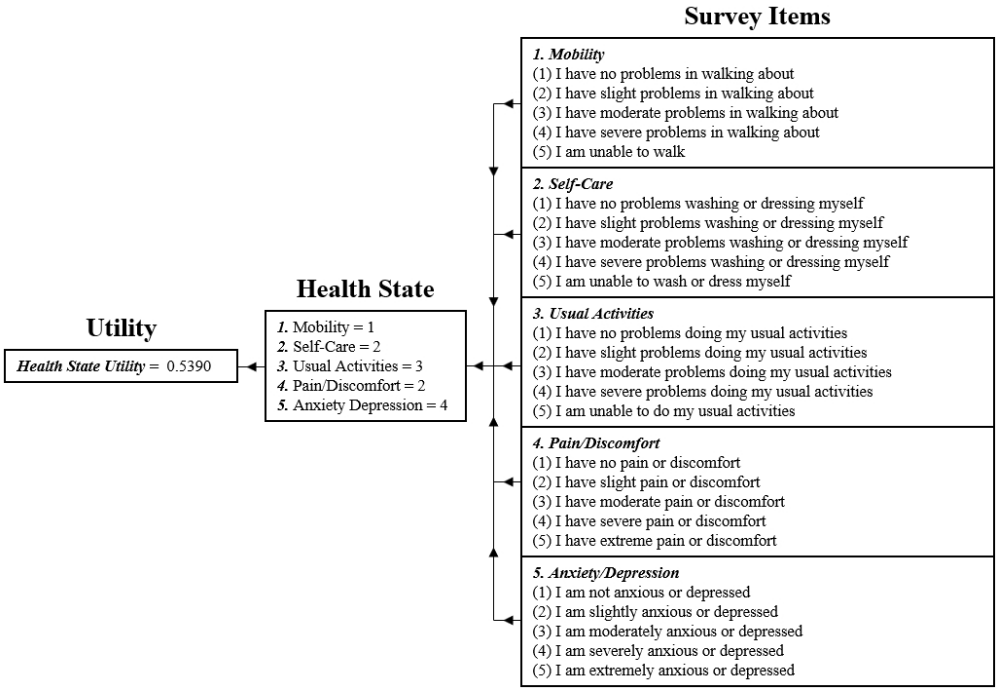
## 6.0 CAPTIONS

### 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.

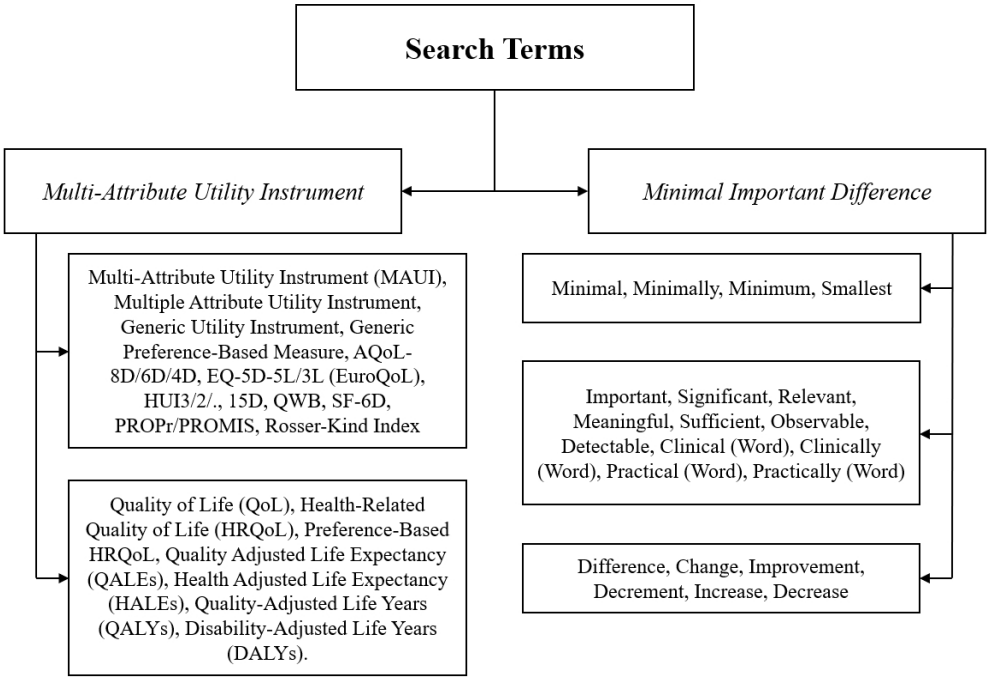
### 6.2 Figure 2 Caption

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'. Additionally, '(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 'Multi-Attribute Utility Instrument' are divided into (top) instrument names (generic and specific) and (bottom) outcome measures associated with multi-attribute utility instruments.




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.

555x384mm (47 x 47 DPI)




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'. Additionally, '(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 'Multi-Attribute Utility Instrument' are divided into (top) instrument names (generic and specific) and (bottom) outcome measures associated with multi-attribute utility instruments.

630x431mm (47 x 47 DPI)

# APPENDIX – SEARCH STRATEGY

*Preface*

The below search strategy will be applied to all databases. The following operators are to be applied in searches:

- Closed captions: used to retrieve exact matches to search phrases.
- Truncation operators: represented below by asterisks, allowed for suffixes and pluralisation of search terms.
- Wildcard operators: represented below by hashes, allowed for variable characters.

The only limit to be imposed on searches involves publication date, with only studies published between 1989 and the present being retrieved in searches. Additionally, Group 1 and Group 2 search terms, listed below, were combined using an 'AND' operator.

*Group 1. Words and phrases associated with "multi-attribute utility instrument"*

"Multi-Attribute Utility Instrument*" or "Multi Attribute Utility Instrument*" or "Multiattribute Utility Instrument*" or "Multiple-Attribute Utility Instrument" or "Multiple Attribute Utility Instrument" or "MAUI*" or "Generic Utility Instrument*" or "Generic Preference-Based Measure*" "Generic Preference Based Measure*" or "AQoL-#D" or "Assessment Quality of Life" or "Assessment Quality-of-Life" or "EuroQoL" or "EQ-#D" or "HUI#" or "Health Utilities Index" or "QWB" or "Quality of Wellbeing" or "Quality of Well-Being" or "15-D" or "15 Dimension*" or "15-Dimension*" or "SF-6D" or "SF-12" or "SF-36" or "Short Form" or "Short-Form" or "EQ-5D-5L Psychosocial" or "EQ-5D-5L-Psychosocial" or "PROMIS" or "Patient-Reported Outcome Measurement Information System" or "Patient Reported Outcome Measurement Information System" or "Rosser-Kind Index" or "PROPr" or "PROMIS Preference" or "Quality of Life" or "Quality-of-Life" or "Preference-Based Health-Related Quality-of-Life" or "Preference Based Health Related Quality of Life" or "Preference-Based Health Related Quality of Life" or "Preference Based Health-Related Quality of Life" or "Preference Based Health Related Quality-of-Life" or "Preference-Based Health-Related Quality of Life" or "Preference Based Health-Related Quality-of-Life" or "Preference-Based Health Related Quality-of-Life" or "Health-Related Quality-of-Life" or "Health-Related Quality of Life" or "Health Related Quality-of-Life" or "Health Related Quality of Life" or "Disability Adjusted Life Years" or "Disability-Adjusted Life Years" or "Disability-Adjusted Life-Years" or "Quality Adjusted Life Years" or "Quality-Adjusted Life Years" or "Quality-Adjusted Life-Years" or "Health Adjusted Life Expectancy" or "Health-Adjusted Life Expectancy" or "Health-Adjusted Life-Expectancy" or "Quality Adjusted Life Expectancy" or Quality-Adjusted Life Expectancy" or Quality-Adjusted Life-Expectancy" or QALY*" or "DALY*" or "HALE*" or "QALE".

*Group 2. Phrases associated with "minimal important difference"*

"Minimum Important Difference*" or "Minimum Significant Difference*" or "Minimum Observable Difference*" or "Minimum Meaningful Difference*" or "Minimum Clinical* Important Difference*" or "Minimum Clinical* Significant Difference*" or "Minimum Clinical* Meaningful Difference*" or "Minimum Clinical* Observable Difference*" or "Minimum Practical* Meaningful Difference*" or "Minimum Practical* Observable Difference*" or "Minimum Practical* Significant Difference*" or "Minimum Practical* Important Difference*" or "Minimal* Important Difference*" or "Minimal* Significant Difference*" or "Minimal* Observable Difference*" or "Minimal* Meaningful Difference*" or "Minimal* Clinical* Important Difference*" or "Minimal* Clinical* Significant Difference*" or "Minimal* Clinical* Meaningful Difference*" or "Minimal* Clinical* Observable Difference*" or "Minimal* Practical* Important Difference*" or "Minimal* Practical* Significant Difference*" or "Minimal* Practical* Observable Difference*" or "Minimal* Practical* Meaningful Difference*" or "Smallest Important Difference*" or "Smallest Significant Difference*" or "Smallest Observable Difference*" or "Smallest Meaningful Difference*" or "Smallest Clinical* Important Difference*" or "Smallest Clinical* Significant Difference*" or "Smallest Clinical* Meaningful Difference*" or "Smallest Clinical* Observable Difference*" or "Smallest Practical* Observable Difference*" or "Smallest Practical* Significant Difference*" or "Smallest Practical* or Important Difference*" or "Smallest Practical* Meaningful Difference*" or "Minimum Relevant Difference*" or "Minimum Clinical* Relevant Difference*" or "Minimum Practical* Relevant Difference*" or "Minimal* Relevant Difference*" or "Minimal* Clinical* Relevant Difference*" or "Minimal* Practical* Relevant Difference*" or "Smallest Relevant Difference*" or "Smallest Clinical* Relevant Difference*" or "Smallest Practical* Relevant Difference*" or "Minimum Important Change*" or "Minimum Significant Change*" or "Minimum Observable Change*" or "Minimum Meaningful Change*" or "Minimum Clinical* Change*" or "Minimum Clinical* Significant Change*" or "Minimum Clinical* Meaningful Change*" or "Minimum Clinical* Observable Change*" or "Minimum Practical* Meaningful Change*" or "Minimum Practical* Observable Change*" or "Minimum Practical* Significant Change*" or "Minimum Practical* Important Change*" or "Minimal* Important Change*" or "Minimal* Significant Change*" or "Minimal* Observable Change*" or "Minimal* Meaningful Change*" or "Minimal* Clinical* Important Change*" or "Minimal* Clinical* Significant Change*" or "Minimal* Clinical* Meaningful Change*" or "Minimal* Clinical* Observable Change*" or "Minimal* Practical* Important Change*" or "Minimal* Practical* Significant Change*" or "Minimal* Practical* Observable Change*" or "Minimal* Practical* Meaningful Change*" or "Smallest Important Change*" or "Smallest Significant Change*" or "Smallest Observable Change*" or "Smallest Meaningful Change*" or "Smallest Clinical* Important Change*" or "Smallest Clinical* Significant Change*" or "Smallest Clinical* Meaningful Change*" or "Smallest Clinical* Observable Change*" or "Smallest Practical* Observable Change*" or "Smallest Practical* Significant Change*" or "Smallest Practical* Important Change*" or "Smallest Practical* Meaningful Change*" or "Minimum Relevant Change*" or "Minimum Clinical* Relevant Change*" or "Minimum Practical* Relevant Change*" or "Minimal* Relevant Change*" or "Minimal* Clinical* Relevant Change*" or "Minimal* Practical* Relevant Change*" or "Smallest Relevant Change*" or "Smallest Clinical* Relevant Change*" or "Smallest Practical* Relevant Change*" or "Minimum Important Improvement*" or "Minimum Significant Improvement*" or "Minimum Observable Improvement*" or "Minimum Meaningful Improvement*" or "Minimum Clinical* Improvement*" or "Minimum Clinical* Significant Improvement*" or "Minimum Clinical* Meaningful Improvement*" or "Minimum Clinical* Observable Improvement*" or "Minimum Practical* Meaningful Improvement*" or "Minimum Practical* Observable Improvement*" or "Minimum

Practical* Significant Improvement*" or "Minimum Practical* Important Improvement*" or "Minimal* Important Improvement*" or "Minimal* Significant Improvement*" or "Minimal* Observable Improvement*" or "Minimal* Meaningful Improvement*" or "Minimal* Clinical* Important Improvement*" or "Minimal* Clinical* Significant Improvement*" or "Minimal* Clinical* Meaningful Improvement*" or "Minimal* Clinical* Observable Improvement*" or "Minimal* Practical* Important Improvement*" or "Minimal* Practical* Significant Improvement*" or "Minimal* Practical* Observable Improvement*" or "Minimal* Practical* Meaningful Improvement*" or "Smallest Important Improvement*" or "Smallest Significant Improvement*" or "Smallest Observable Improvement*" or "Smallest Meaningful Improvement*" or "Smallest Clinical* Important Improvement*" or "Smallest Clinical* Significant Improvement*" or "Smallest Clinical* Meaningful Improvement*" or "Smallest Clinical* Observable Improvement*" or "Smallest Practical* Observable Improvement*" or "Smallest Practical* Significant Improvement*" or "Smallest Practical* Important Improvement*" or "Smallest Practical* Meaningful Improvement*" or "Minimum Relevant Improvement*" or "Minimum Clinical* Relevant Improvement*" or "Minimum Practical* Relevant Improvement*" or "Minimal* Relevant Improvement*" or "Minimal* Clinical* Relevant Improvement*" or "Minimal* Practical* Relevant Improvement*" or "Smallest Relevant Improvement*" or "Smallest Clinical* Relevant Improvement*" or "Smallest Practical* Relevant Improvement*" or "Minimum Important Decrement*" or "Minimum Significant Decrement*" or "Minimum Observable Decrement*" or "Minimum Meaningful Decrement*" or "Minimum Clinical* Decrement*" or "Minimum Clinical* Significant Decrement*" or "Minimum Clinical* Meaningful Decrement*" or "Minimum Clinical* Observable Decrement*" or "Minimum Practical* Meaningful Decrement*" or "Minimum Practical* Observable Decrement*" or "Minimum Practical* Significant Decrement*" or "Minimum Practical* Important Decrement*" or "Minimal* Important Decrement*" or "Minimal* Significant Decrement*" or "Minimal* Observable Decrement*" or "Minimal* Meaningful Decrement*" or "Minimal* Clinical* Important Decrement*" or "Minimal* Clinical* Significant Decrement*" or "Minimal* Clinical* Meaningful Decrement*" or "Minimal* Clinical* Observable Decrement*" or "Minimal* Practical* Important Decrement*" or "Minimal* Practical* Significant Decrement*" or "Minimal* Practical* Observable Decrement*" or "Minimal* Practical* Meaningful Decrement*" or "Smallest Important Decrement*" or "Smallest Significant Decrement*" or "Smallest Observable Decrement*" or "Smallest Meaningful Decrement*" or "Smallest Clinical* Important Decrement*" or "Smallest Clinical* Significant Decrement*" or "Smallest Clinical* Meaningful Decrement*" or "Smallest Clinical* Observable Decrement*" or "Smallest Practical* Observable Decrement*" or "Smallest Practical* Significant Decrement*" or "Smallest Practical* Important Decrement*" or "Smallest Practical* Meaningful Decrement*" or "Minimum Relevant Decrement*" or "Minimum Clinical* Relevant Decrement*" or "Minimum Practical* Relevant Decrement*" or "Minimal* Relevant Decrement*" or "Minimal* Clinical* Relevant Decrement*" or "Minimal* Practical* Relevant Decrement*" or "Smallest Relevant Decrement*" or "Smallest Clinical* Relevant Decrement*" or "Smallest Practical* Relevant Decrement*" or "Minimum Important Increase*" or "Minimum Significant Increase*" or "Minimum Observable Increase*" or "Minimum Meaningful Increase*" or "Minimum Clinical* Increase*" or "Minimum Clinical* Significant Increase*" or "Minimum Clinical* Meaningful Increase*" or "Minimum Clinical* Observable Increase*" or "Minimum Practical* Meaningful Increase*" or "Minimum Practical* Observable Increase*" or "Minimum Practical* Significant Increase*" or "Minimum Practical* Important Increase*" or "Minimal* Important Increase*" or "Minimal* Significant Increase*" or "Minimal* Observable Increase*" or "Minimal* Meaningful Increase*" or "Minimal* Clinical* Important Increase*" or "Minimal* Clinical* Significant Increase*" or "Minimal* Clinical* Meaningful Increase*" or "Minimal* Clinical* Observable Increase*" or "Minimal* Practical* Important Increase*" or "Minimal* Practical* Significant Increase*" or "Minimal* Practical* Observable Increase*" or "Minimal* Practical* Meaningful Increase*" or "Smallest Important Increase*" or "Smallest Significant Increase*" or "Smallest Observable Increase*" or "Smallest Meaningful Increase*" or "Smallest Clinical* Important Increase*" or "Smallest Clinical* Significant Increase*" or "Smallest Clinical* Meaningful Increase*" or "Smallest Clinical* Observable Increase*" or "Smallest Practical* Observable Increase*" or "Smallest Practical* Significant Increase*" or "Smallest Practical* Important Increase*" or "Smallest Practical* Meaningful Increase*" or "Minimum Relevant Increase*" or "Minimum Clinical* Relevant Increase*" or "Minimum Practical* Relevant Increase*" or "Minimal* Relevant Increase*" or "Minimal* Clinical* Relevant Increase*" or "Minimal* Practical* Relevant Increase*" or "Smallest Relevant Increase*" or "Smallest Clinical* Relevant Increase*" or "Smallest Practical* Relevant Increase*" or "Minimum Important Decrease*" or "Minimum Significant Decrease*" or "Minimum Observable Decrease*" or "Minimum Meaningful Decrease*" or "Minimum Clinical* Decrease*" or "Minimum Clinical* Significant Decrease*" or "Minimum Clinical* Meaningful Decrease*" or "Minimum Clinical* Observable Decrease*" or "Minimum Practical* Meaningful Decrease*" or "Minimum Practical* Observable Decrease*" or "Minimum Practical* Significant Decrease*" or "Minimum Practical* Important Decrease*" or "Minimal* Important Decrease*" or "Minimal* Significant Decrease*" or "Minimal* Observable Decrease*" or "Minimal* Meaningful Decrease*" or "Minimal* Clinical* Important Decrease*" or "Minimal* Clinical* Significant Decrease*" or "Minimal* Clinical* Meaningful Decrease*" or "Minimal* Clinical* Observable Decrease*" or "Minimal* Practical* Important Decrease*" or "Minimal* Practical* Significant Decrease*" or "Minimal* Practical* Observable Decrease*" or "Minimal* Practical* Meaningful Decrease*" or "Smallest Important Decrease*" or "Smallest Significant Decrease*" or "Smallest Observable Decrease*" or "Smallest Meaningful Decrease*" or "Smallest Clinical* Important Decrease*" or "Smallest Clinical* Significant Decrease*" or "Smallest Clinical* Meaningful Decrease*" or "Smallest Clinical* Observable Decrease*" or "Smallest Practical* Observable Decrease*" or "Smallest Practical* Significant Decrease*" or "Smallest Practical* Important Decrease*" or "Smallest Practical* Meaningful Decrease*" or "Minimum Relevant Decrease*" or "Minimum Clinical* Relevant Decrease*" or "Minimum Practical* Relevant Decrease*" or "Minimal* Relevant Decrease*" or "Minimal* Clinical* Relevant Decrease*" or "Minimal* Practical* Relevant Decrease*" or "Smallest Relevant Decrease*" or "Smallest Clinical* Relevant Decrease*" or "Smallest Practical* Relevant Decrease*".

# PRISMA-P 2015 Checklist

| Section/topic | # | Checklist item | Information reported | | Line number(s) |
|---|---|---|---|---|---|
| | | | Yes | No | |
| **ADMINISTRATIVE INFORMATION** | | | | | |
| **Title** | | | | | |
| Identification | 1a | Identify the report as a protocol of a systematic review | X | | Page 4, lines 1-4 |
| Update | 1b | If the protocol is for an update of a previous systematic review, identify as such | NA | | |
| **Registration** | 2 | If registered, provide the name of the registry (e.g., PROSPERO) and registration number in the Abstract | X | | Page 4, lines 3-4 |
| **Authors** | | | | | |
| Contact | 3a | Provide name, institutional affiliation, and e-mail address of all protocol authors; provide physical mailing address of corresponding author | X | | Page 1, lines 3-28 |
| Contributions | 3b | Describe contributions of protocol authors and identify the guarantor of the review | X | | Page 14, lines 2-9 |
| **Amendments** | 4 | If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments | NA | | |
| **Support** | | | | | |
| Sources | 5a | Indicate sources of financial or other support for the review | X | | Page 14, lines 10-12 |
| Sponsor | 5b | Provide name for the review funder and/or sponsor | NA | | |
| Role of sponsor/funder | 5c | Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol | NA | | |
| **INTRODUCTION** | | | | | |
| **Rationale** | 6 | Describe the rationale for the review in the context of what is already known | X | | Pages 8(9), lines 3-12(5-12) |
| **Objectives** | 7 | Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO) | X | | Pages 8(9), lines 13-32(1-4) |
| **METHODS** | | | | | |
| **Eligibility criteria** | 8 | Specify the study characteristics (e.g., PICO, study design, setting, time frame) and report characteristics (e.g., years considered, language, publication status) to be used as criteria for eligibility for the review | X | | Page 11, lines 5-12 |
| **Information sources** | 9 | Describe all intended information sources (e.g., electronic databases, contact with study authors, trial registers, or other grey literature sources) with planned dates of coverage | X | | Pages 10(11), lines 27-32(1-3) |
| **Search strategy** | 10 | Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated | X | | Appendix |

| Section/topic | # | Checklist item | Information reported | | Line number(s) |
|---|---|---|---|---|---|
| | | | Yes | No | |
| ***STUDY RECORDS*** | | | | | |
| Data management | 11a | Describe the mechanism(s) that will be used to manage records and data throughout the review | X | | Page 12, lines 12-16 |
| Selection process | 11b | State the process that will be used for selecting studies (e.g., two independent reviewers) through each phase of the review (i.e., screening, eligibility, and inclusion in meta-analysis) | X | | Page 11, lines 13-21 |
| Data collection process | 11c | Describe planned method of extracting data from reports (e.g., piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators | X | | Page 11, lines 14-15 and 22-25 |
| **Data items** | 12 | List and define all variables for which data will be sought (e.g., PICO items, funding sources), any pre-planned data assumptions and simplifications | X | | Pages 11(12), lines 26-29(1-11) |
| **Outcomes and prioritization** | 13 | List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale | X | | Page 9, lines 5-12 |
| **Risk of bias in individual studies** | 14 | Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis | X | | Pages 9(10), lines 24-30(1-12) |
| ***DATA*** | | | | | |
| **Synthesis** | 15a | Describe criteria under which study data will be quantitatively synthesized | X | | Page 12, lines 23-25 |
| | 15b | If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data, and methods of combining data from studies, including any planned exploration of consistency (e.g., $I^2$, Kendall's tau) | X | | Page 12, lines 17-28 |
| | 15c | Describe any proposed additional analyses (e.g., sensitivity or subgroup analyses, meta-regression) | X | | Pages 12(13), lines 28-30(1-27) |
| | 15d | If quantitative synthesis is not appropriate, describe the type of summary planned | NA | | |
| **Meta-bias(es)** | 16 | Specify any planned assessment of meta-bias(es) (e.g., publication bias across studies, selective reporting within studies) | X | | Page 10, lines 1-10 |
| **Confidence in cumulative evidence** | 17 | Describe how the strength of the body of evidence will be assessed (e.g., GRADE) | X | | Page 10, lines 10-12 |