Should minimally important differences be applied to EQ-5D instruments?

Author: Giselle Abangma, Department of Health Services Research & Policy, London School of Hygiene & Tropical Medicine, UK

Co-Authors: Andrew Briggs, Department of Health Services Research & Policy, London School of Hygiene & Tropical Medicine, UK

Andrew Lloyd, Acaster Lloyd Consulting Ltd. UK

David G T Whitehurst, Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada

Abstract

Background: EQ-5D instruments are one of the most frequently used generic, preference-based health-related quality of life instruments (HRQoL). Several studies have investigated whether observed differences in the index scores of EQ-5D instruments are meaningful to patients using the concept of minimally important difference (MID). There is currently no consensus on whether MID is a useful concept to apply to cardinal utility measures such as EQ-5D instruments that are used in economic evaluation.

Objective: To critically evaluate the use of MID for EQ-5D instruments and for QALY-based health economic evaluation.

Methods: A scoping review was conducted using Arksey and O'Malley's methodological framework and followed PRISMA recommendations for scoping reviews. We searched PubMed and Google scholar using selected keywords. We only included studies that used generic preference-based HRQoL instruments and excluded non-English language studies. We explored how the MID for EQ-5D instruments have been used in economic evaluation and offer critical comment on the role of MIDs in economic evaluation, particularly the methods for calculating MID.

Results: MID for EQ-5D instruments have been estimated across many disease areas, using several different methodologies. These reported MIDs vary widely and lack consistency. No study reported on how MID for EQ-5D instruments can be used for economic evaluation though they have been used in cost minimisation analysis for some non-inferiority trials and in sample size calculations for economic evaluations done alongside clinical trials. Some of the issues identified as to why no study reported on how MID for EQ-5D instruments can be used for economic evaluation include: the way in which the EQ-5D instruments index scores are derived; the influence of country-specific weights on the estimates; the methods used for estimation; the impact of individual MID on population MID; and the usefulness of calculating MIDs for generic cardinal utility measures like the EQ-5D instruments.

Conclusions: The concept of the MID has become well accepted in clinical research, particularly in the area of disease-specific patient reported outcomes. Although methods used in this area can also be used to calculate the MID for generic HRQoL instruments such as the EQ-5D instruments, it is less clear that the MID concept should be used with cardinal utility measures and for cost-effectiveness studies. As a result, we consider the concept is currently over-used and over-reported.

Introduction

EQ-5D instruments are the most widely used measure of health-related quality of life (HRQoL) around the world. They are a generic measure of HRQoL accompanied by value sets used in the calculation of QALYs (quality adjusted life years). EQ-5D instruments have become the most used multi attribute utility instrument for measuring HRQoL in cost-effectiveness analysis [1] and health technology appraisal (HTA). Several studies [2-12] have investigated whether observed changes are meaningful to patients using the concept of minimally important difference (MID). The MID is defined as the smallest change in an outcome that a patient would identify as important [13]. There are several MID-related terminologies in the literature such as minimal clinically important difference (MCID), clinically important difference, minimally detectable difference (MDD), and meaningful change threshold (MCT) with very subtle differences in definition between them. However, as King [14] points out, these differences in terminology have little impact on the way these quantities are estimated and used. The terminology has evolved over time and MID has become the dominant term in the literature.

The concept of the MID originated in clinical research to support the interpretation of patient reported outcomes (PROs). Over the last two decades, there has been increasing focus on incorporating the patient perspective in the interpretation of treatment outcomes [15, 16]. As such PRO instruments have become an integral part of determining the effectiveness of a treatment. In clinical research, statistical significance is used to estimate the difference in outcomes scores between two treatment strategies or trial arms. However, when statistically significant differences are observed, it may remain unclear whether these differences are clinically relevant or not, or indeed whether the difference reflects an important difference for patients.

The interpretation of changes in PRO scores is challenging largely because non-preference weighted measures don't have a common measurement scale. The response options and the scores from PRO scales are ordinal with no true zero value. This makes it difficult to draw robust conclusions regarding changes in HRQoL and determine how meaningful a change on the scale is to the patient, and whether the change has any clinical relevance. For example, the Problem Areas In Diabetes (PAID) questionnaire, sometimes referred to as the PAID-20, is a questionnaire with 20 items and the following five response options: 1 = not a problem, 2 =

minor problem, 3 = moderate problem, 4 = somewhat serious problem, 5 = serious problem. The scores for each item are summed, then multiplied by 1.25 to generate a total score out of 100 [17]. The total score has no interval properties i.e. a change from 30 to 50 may not be twice as good as a change from 10 to 20. Similarly, the difference between 10 and 30 may not mean the same as the difference between 20 and 40. Recognition of this serious limitation led to increasing interest in the use of the concept of MID [18]. The MID places the magnitude of change in context to help clinicians assess whether interventions result in meaningful improvement in patients' HRQoL and other PRO scores. As a result, the MID has become a standard approach in the interpretation of clinical relevance of changes in PROs [19].

This practice has spilled over to the EQ-5D instruments and other preference-based instruments, with MIDs calculated for various disease areas across many populations using varying methodologies resulting in varied MID values [3, 5, 10, 12, 20-22]. One reason for this is because EQ-5D instruments are often used as clinical endpoints by researchers with no interest in cost effectiveness analysis.

The calculation of a MID for generic preference-based HRQoL instruments such as the EQ-5D instruments is a topic on which there is currently no consensus, either to its usefulness or the best methods for its estimation. We aim to contribute to providing some clarity by 1) providing a description of methods used for estimating MIDs for HRQoL instruments and other PRO measures. 2) systematically scoping the literature to understand how the MID for EQ-5D instruments has been used in economic evaluation. 3) discussing reasons why the MID concept should not be applied to EQ-5D instruments and other preference-based measures. 4) providing a critique of the instrument-defined approach for determining MID for EQ-5D instruments and other preference-based instruments.

Methods for estimating MIDs for HRQoL instruments and other PRO measures.

There are mainly two types of methods for estimating the MID - the anchor-based approach and the distribution-based approach [14, 19, 23, 24].

Anchor-based approaches examine the relationship between a HRQoL measure (and other PROs) with another measure of change - the anchor[25]. The anchor-based approach uses a validated external indicator to classify patients as improved or worsened and estimates the degree of change in the target measure as the difference between those that change and those who are stable. A global assessment tool where patients classify themselves as unchanged, or experiencing small, moderate, and large improvement or deterioration is usually used as an external criterion to distinguish meaningful changes in scores. The MID is estimated as the mean difference in the PRO score that is derived from patients in the small change groups. Numerical rating scales are also used where everyone who has the same score at two time points is considered stable, and people with different scores are considered changed. It is generally agreed that the patient-reported anchor-based approach is the optimal way to determine the MID because scores are rated against people's perception of change [14, 18, 24, 26, 27]. A limitation of this approach is that the MID is dependent on the type of anchor used and on an assumption that the anchor correctly distinguishes between important and unimportant changes in health states [28].

With the distribution-based approach, the MID is defined based on statistical distributions of HRQoL scores. The approach relies on relating the difference between treatment and control groups or change from baseline to some measure of variability such as the standard error of measurement (SEM) and effect size (ES) amongst others [29]. The SEM is the variation in scores attributed to instrument unreliability. To define the MID, threshold values of 1 SEM have been suggested in the literature [30-32] hence changes smaller than 1 SEM are likely due to measurement error rather than a true observed change. The ES is a standardized measure of change. It is the ratio of mean change and the SD of the baseline values. A fifth and a half of a SD are two effect size summary statistics commonly used to estimate the MID [20, 33, 34] Cohen's d effect sizes of 0.2, 0.5, and 0.8 are commonly used to interpret changes as small, medium, and large effect sizes, respectively [35, 36] and used as cut off points to define the MID. The change in scores corresponding to the small effect size is considered the MID. A drawback of this approach is that it does not include the patient's own evaluation, instead relies solely on the statistical properties of the measure and provides no information on the very thing it is trying to measure – the importance of changes to the patient.

Although there's no consensus as to which is the best method for estimating the MID, the general recommendation has been to use anchor-based methods with triangulation of findings from multiple anchors and other approaches where possible [26, 37].

Methodology

A scoping review was conducted to examine how the MID for EQ-5D instruments have been used in cost-effectiveness analysis. Unlike systematic reviews, the aim of a scoping review is not to systematically identify evidence that is relevant to a particular question but to map the field [38]. Hence, scoping reviews have been referred to as gaining an understanding of the "lay of the land". Scoping reviews are useful for examining the extent, range and nature of a research area and identifying available literature on a topic particularly when a research area has not yet been widely reviewed as is the case with MIDs for EQ-5D instruments. It aims to rapidly map the key concepts underpinning a research area especially where there is a lack of consistency in methodology and terminology and summarises the main sources and types of evidence available [39-41]. It is also useful in informing if a full systematic reviews is needed [38]. For these reasons, a scoping review was utilised as it met the objective of the review question.

The scoping review was conducted using Arksey and O'Malley's methodological framework [41] featuring refinements and recommendations made by the Joanna Briggs Institute [42]. The PRISMA recommendations for scoping reviews was also followed. The following steps were followed: 1) identifying the research question; 2) identifying relevant studies; 3) study selection; 4) charting the data; and 5) collating, summarising, and reporting the results.

Identifying the research question

The research question posed in this review is to evaluate the use of MID for EQ-5D instruments and for economic evaluation.

Identifying relevant studies.

We searched PubMed and Google scholar using variations of the following key words ("minimally important difference", "Health related quality of life", "Preference-based instrument", "EQ-5D", "SF-6D", "HUI", "Assessment of Quality of Life", "15D ", and "economic evaluation"). The search was extended to include not just EQ-5D instruments but other preference-based HRQoL instruments used in economic evaluation for completeness. All bibliographies of included studies were manually checked for additional references. We searched for studies written in English with no year restriction.

Study selection

Studies were eligible for inclusion if they used MID for EQ-5D instruments or other preference-based HRQoL instruments in economic evaluation. Titles and abstracts of identified papers were screened for potential eligibility after downloading the search results into Endnote. The full texts of potentially eligible articles were retrieved and read to determine eligibility for final inclusion.

Charting the data

A data charting form was developed in Microsoft Excel. The following information were extracted from each individual paper: (a) basic study characteristics (i.e. authors, title, study year); (b) disease area; (c) estimate calculated (i.e. cost/MID, ICER/MID); (d) type of preference-based instrument used; and (e) rationale for using MID in economic evaluation

Collating, summarising, and reporting the results

The extracted data were analysed and examined in detail and presented in a table.

Results

Seventy-one articles were identified in total. Sixty-nine of which were identified via PubMed and 2 via Google scholar. Fifty-seven studies were excluded as they were not relevant to the aim of the study because they did not involve the use of MID for EQ-5D instruments or other preference-based instruments in economic evaluation. Eight articles were excluded as the MID wasn't mentioned. Six studies where identified that used MIDs derived from disease-specific instruments for economic evaluation. No study was identified that used the MID for EQ-5D instruments or any other generic preference-based HRQoL instrument in economic evaluation. The flow chart of study selection is presented in Fig. 1.

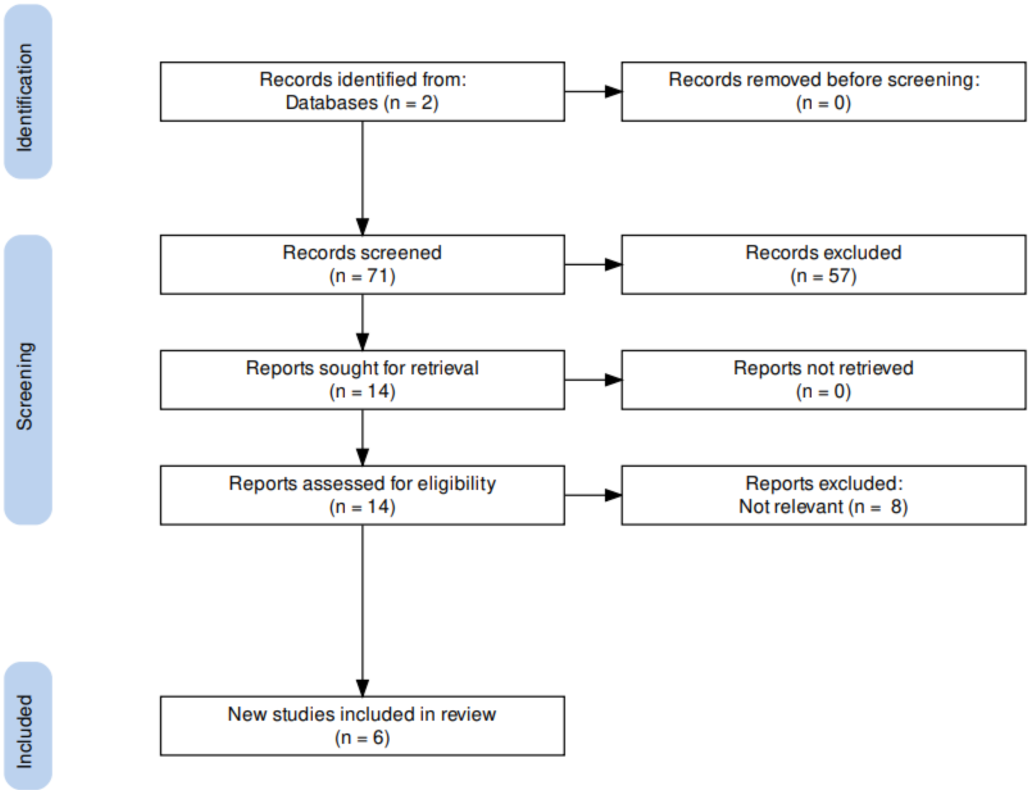

Fig. 1 PRISMA flow chart of study selection

The review showed that the MID for EQ-5D instruments have been estimated and used across many disease areas with heterogeneity in the estimates. In general, the MID for EQ-5D instruments was calculated or used for the following reasons: 1) for sample size calculation; 2) to study the psychometric properties of EQ-5D instruments, including establishing a MID; 3) as effect size in outcome measures; 4) used to estimate MID for a particular disease area and population; 5) used to establish the threshold value (margin) at which noninferiority or equivalence can be declared. Given that the focus of the search was the use of MID for EQ-5D instruments in economic evaluation, the identified studies are likely to be a smaller subset of the literature on the use of MIDs for EQ-5D instruments.

Of the 6 studies [43-48] that used MIDs derived from disease-specific instruments for economic evaluation (see table 1), the MIDs were used to estimate the costs/MID improvement (responder analysis) and used to calculate incremental cost-effectiveness ratio (ICER) per MID improvement (including summarising uncertainty around the ICER using cost-effectiveness acceptability curves of the MID). Responder definitions focus on individual level change and assist in understanding proportion of patients benefiting from treatment by simply counting the number of patients who improved by more than the chosen response threshold. As such, a responder is a patient who has experienced a change that is important to that patient. Of the 6 studies, 3 also calculated ICER/QALY using the EQ-5D.

No articles were identified that looked at the methodological aspect of calculating the MID for preference-based HRQoL measures such as the EQ-5D instruments. However, 1 study [49] looked at the MID concept more generally in HRQoL research (not specific to EQ-5D instruments) and identified several practical problems in estimating MIDs including the fact that the estimated magnitude varies depending on the distributional index and the external standard or anchor; the amount of change might depend on the direction of change; and the meaning of change depends on where you start (baseline value). The study recommended that caution be taken when using the MID.

Table 1 Summary of studies that utilised MIDs derived from disease-specific instruments in cost-effectiveness analysis

Study	Disease area	Estimate	Instrument used	Rationale for using MID
Fueki kenji et al, 2021[45]	Dental	ICER /MID improvement	Oral health-related quality of life (OHRQoL)	Main treatment effect
Goldstein et al, 1997[46]	Respiratory function	ICER/MID improvement	Abstract only	Abstract only
McKenna et al, 2014[47]	Dental	Cost effectiveness ratio	Oral health-related quality of life (OHRQoL)	Main treatment effect
Verberkt et al, 2021[48]	COPD	ICER/MID improvement ICER/QALY*	COPD Assessment Test (CAT); EQ-5D-5L	Disease-specific HRQoL measure
Böckmann et al, 2021[43]	Asthma	ICER/MID improvement ICER/QALY*	Asthma Control Test , Asthma Quality of Life Questionnaire; EQ-5D-5L	To include patients' perspective
Brusco et al, 2015[44]	Rehabilitation	ICER/MID improvement ICER/QALY*	Functional independence measure (FIM); EQ-5D-3L	No rationale provided

* These studies used both an EQ-5D instrument and a disease-specific instrument. However, MID for EQ-5D instruments were not used for cost effectiveness analysis. The MIDs were derived using the corresponding disease-specific measure in the table.

Discussion

The review showed that the MID for EQ-5D instruments vary widely and lack consistency. No study reported on how MID for EQ-5D instruments can be used for economic evaluation and no articles were identified that explored the methodological challenges in calculating MID for the EQ-5D instruments. Here we present some arguments for discussion as to why the MID should not be estimated for EQ-5D instruments and used for QALY based economic evaluation. We also provide a critique of the instrument-defined approach used for calculating MIDs for preference-based HRQoL instruments. Just because we can estimate the MID for EQ-5D instruments doesn't mean we should.

Costs are not considered.

Studies estimating the MID for EQ-5D instruments do not consider the associated costs. Clinical trials measure health outcomes to determine the efficacy of a health technology. If resources were unlimited, clinical efficacy and safety would be the only criteria to consider when making allocation and implementation decisions. If there were no risks and no costs, any improvement in HRQoL would be worth having no matter how small [32]. Given that resources are limited it is imperative we know which technology is the most cost-effective option. To determine if an intervention is cost-effective, the estimation of the joint density of costs and effects differences of competing interventions need to be considered [50], therefore considering the effects of a treatment alone is meaningless in cost-effectiveness analysis [51]. Hence, the changes in health outcomes attributed to a new intervention must be compared with the associated cost to inform allocation and implementation decisions. Similarly, the opportunity cost must be considered. Not considering costs could result in cost-effective interventions not being implemented. For example, an intervention with an incremental cost-effectiveness ratio of say £1000/QALY could be rejected because it falls below the MID threshold.

Calculation of index scores for EQ-5D instruments

The way values are calculated for EQ-5D instruments makes the determination of the MID irrelevant. Calculation of the index value includes the use of value sets which reflects how good or bad a health state is according to the preferences of the general population of a country/region. As such the calculated MID embodies the preferences of the general population and is not solely a reflection of what the patient deems as important. Additionally, the calculated MID will depend on the value set used as each value set places a different weight on the various levels and dimensions of the profile data, reflecting underlying differences in preferences. For example, a study that estimated and compared the MID in index score of country-specific EQ-5D-5L scoring algorithms for Germany, Indonesia, Ireland, Malaysia, Poland, Portugal, Taiwan, and the United States using the instrument-defined approach reported mean estimates ranging from 0.072 for the Malaysian tariff to 0.101 for Taiwan, with German, Indonesian, Irish, Polish, Portuguese, and US values falling between these bounds [7]. Another study reported MID values of 0.056, 0.069, 0.061, 0.048, 0.063 and 0.063 for Canada, China, Spain, Japan, England, and Uruguay respectively for instrument-defined MID estimates for EQ-5D-5L country-specific scoring algorithms [3].

When the MID is estimated using psychometric measures, it is usually very specific to a concept or construct for example pain or depression. In these instances, conclusions can be made as to whether the patient has achieved an improvement in their pain or not. This is not so straightforward for index scores of EQ-5D instruments given that it is a single index measure that summarises 5 dimensions of health. As such, the score isn't for a single construct but for a combination of 5 dimensions. For example, a person's overall utility score could (in principle) go up even though some aspects of their health are getting worse. If dimensions of health are changing in different directions, then it seems hard to consider this an MID shift.

Utilities are measured on a cardinal scale of 0-1, where 0 indicates dead and 1 indicates full health. Using the 'anchors' of 0 and 1, utility measurement is on an interval scale, where the same change means the same irrespective of the part of the scale being considered (e.g., a change in health from 0.1 to 0.2 is equivalent to a change from 0.7 to 0.8). States worse than death have negative values. Given the presence of a true zero and the fact that health state utility values reflect the desirability or preference for a particular health state compared to another and by how much suggests that there is no conceptual basis for a MID for the values of EQ-5D instruments.

The design of EQ-5D instruments

Extensive research has shown the EQ-5D instruments to be valid, reliable, and responsive in a wide range of conditions and populations [1, 11, 52-56]. For a valid, reliable, and responsive multi attribute health classification system, any movement on the instrument should be considered an important change - not just those exceeding the MID - since the instrument is designed to describe only meaningful differences or changes in health status and not trivial or unimportant ones. As a result, any change in health states should be considered an important difference.

Group level MIDs cannot be used as responder thresholds for individual-level change.

As stated previously, the MID is used for interpreting mean group differences between treatments [24, 57]. A group level MID is not suitable for identifying patients that benefit from treatment (responders). The goal of estimating the MID is to identify an absolute minimum that must be exceeded before changes are considered worthwhile, changes below the MID are considered worthless. Guyatt, recommends interpreting results in ways that consider the proportion of patients achieving the incremental benefit may be more important than simply comparing mean differences[24]. This is where responder definition is used. A responder is a patient who has experienced a change that is important to them. Responder definitions focus on the individual and helps in understanding the number of patients benefiting from treatment. Identifying those who improve ("responders" to treatment) provides important additional information to group changes. Hays et al [58] explain that using group level MID to identify responders would lead to wrongful classification of patients as responders when they're not because a bigger change is needed for statistically significant change in an individual's score compared to a group change due to larger standard errors for estimates of individual change.

A critique of the instrument-defined approach – a method for estimating MID for preference-based HRQoL measures.

The instrument-defined approach was first described by Luo et al [12] as a method to estimate the MID for preference-based HRQoL instruments based on the health classification system and the utility function. This method is not applicable to psychometrically scored HRQoL instruments since they are usually scored on multiple items with equal weights. Luo et al estimate the MID by taking an average of all the smallest health transitions described by the health classification system. They define the smallest health transitions as transitions in any adjacent pair of health states that differ in only 1 health dimension and by only one level, holding all other dimensions constant. So, for EQ-5D instruments, for example, health states of 21111 and 21112. Each health transition provides 1 estimate of the MID, which is calculated by subtracting the lower index score from the higher index score of the 2 health states defining the transition. However, Luo et al [12] explain that it is possible that certain single-level transitions result in changes in the EQ-5D instrument index score that may be considered larger than the MID, for example, transitions between levels 2 and 3 may be larger than transitions between levels 1 and 2 within a dimension of the EQ-5D-3L as illustrated here - the EQ-5D-3L value for transitions between health states 22221 and 32221 (0.358) is almost 6 times higher than the change in EQ-5D-3L value for the transition between health states 12221 and 22221 (0.060) using the US value set [12]. The largest shift in index score resulting from a single-level transition within each dimension are referred to in the literature as "maximum-valued scoring parameters"[3, 7]. In a study by McClure and colleagues, they showed that exclusion of maximum-valued scoring parameters results in different MID estimates [3]. The results were as follows: Canada, 0.056; China, 0.069; Spain, 0.061; Japan, 0.048; England, 0.063; and Uruguay, 0.063. After excluding the maximum-valued scoring parameters, the MID estimates were as follows: Canada, 0.037; China, 0.058; Spain, 0.045; Japan, 0.044; England, 0.037; and Uruguay, 0.040. Methodological limitations identified by Luo et al include the fact that many of the health transitions included in estimating the MID might never occur in reality and inclusion of these transitions might lead to biased estimates. However, analyses of large datasets suggests that almost all health states are recorded – albeit some are very infrequent. Additionally, they identified that some health transitions used in estimating the MID may represent trivial or large changes that are not suitable for MID estimation and their inclusion may again lead to biased estimates. We argue that their classification of transitions as trivial is based on arbitrary judgement. How is the cut off value for trivial transitions determined?

Luo et al and other commentators [3, 7, 12] have classified the instrument-defined approach as a variant of the anchor-based approach where the instrument serves as an internal anchor. For the purposes of this paper, we consider it a stand-alone method and not a variant of the anchor-based approach because for an anchor-based approach you need an external independent anchor whereas as Luo et al describes it, the instrument-defined approach uses the functional levels of health dimensions embedded in the health classification system as internal anchors with no external independent anchors.

The instrument-defined approach is based on the assumption that any difference in levels within any dimension of the health classification system represents an important difference in health with a single-level transition representing an MID [3, 7, 12]. By this definition, we propose that the MID should be estimated simply by choosing the smallest of all single-level transitions as we believe this approach would be conceptually more consistent. By our definition, the 3-level version would produce larger MID estimates compared to the 5 level more sensitive version. Mulhern et al [59] compared differences in utility between adjacent states using 3 different UK value sets and reported that comparisons of the matched states demonstrate that the change in adjacent states is substantially larger for the three-level tariff across all five dimensions compared to the five-level. This is expected since the EQ-5D-3L includes three severity levels (none, some, extreme/unable to) and describes 243 health states while the five-level version includes five response levels (none, slight, moderate, severe, extreme/unable to) and describes 3125 health states. As a result, the three-level produces larger MIDs compared to the five-level. An instrument-defined MID estimate for the EQ-5D-3L of 0.082 has been reported for the UK scoring algorithm[12] which is higher compared to an instrument-defined MID estimate of 0.037 for EQ-5D-5L [3]. We recognise that this proposed method may simply reflect the smallest difference that the instrument can measure rather than a MID and therefore actually reflects an upper bound to the MID. Additionally, some might argue that the calculated MID does not reflect the smallest score that people find important, but the smallest difference between the health state descriptions, which is fixed by the descriptive system itself, not by the people who value them [28]. However, if the EQ-5D instruments have been proven to be sensitive and responsive, then by design they should only be describing meaningful changes in health.

Conclusion

The concept of the MID has become well accepted in clinical research, particularly in disease-specific PROs. For reasons discussed in this paper, we believe the MID concept should not be applied to EQ-5D instruments and other cardinal utility measures or used for QALY-based health economic evaluation. As a result, we consider the concept is currently over-used and over-reported. At the very least, authors presenting MIDs for EQ-5D instruments should be justifying why MID is an important concept for EQ-5D instruments and how they anticipate the results of their analyses might inform health economic evaluation.

References

1. EuroQol. [cited 2022 19 APR 2022]; Available from: https://euroqol.org/eq-5d-instruments/.
2. McClure, N.S., et al., *Minimally important difference of the EQ-5D-5L index score in adults with type 2 diabetes.* Value in Health, 2018. **21**(9): p. 1090-1097.
3. McClure, N.S., et al., *Instrument-defined estimates of the minimally important difference for EQ-5D-5L index scores.* Value in Health, 2017. **20**(4): p. 644-650.
4. Tsai, A.P.Y., et al., *Minimum important difference of the EQ-5D-5L and EQ-VAS in fibrotic interstitial lung disease.* Thorax, 2021. **76**(1): p. 37-43.
5. Pickard, A.S., M.P. Neary, and D. Cella, *Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer.* Health and quality of life outcomes, 2007. **5**(1): p. 1-8.
6. Xu, R.H., E.L. Wong, and A.W. Cheung, *Estimation of minimally important difference of the EQ-5D-5L utility scores among patients with either hypertension or diabetes or both: a cross-sectional study in Hong Kong.* BMJ Open, 2020. **10**(11): p. e039397.
7. Henry, E.B., et al., *Estimation of an instrument-defined minimally important difference in EQ-5D-5L index scores based on scoring algorithms derived using the EQ-VT version 2 valuation protocols.* Value in Health, 2020. **23**(7): p. 936-944.
8. Kim, S.K., et al., *Estimation of minimally important differences in the EQ-5D and SF-6D indices and their utility in stroke.* Health Qual Life Outcomes, 2015. **13**: p. 32.
9. Szentes, B.L., et al., *How does the EQ-5D-5L perform in asthma patients compared with an asthma-specific quality of life questionnaire?* BMC Pulm Med, 2020. **20**(1): p. 168.
10. Walters, S.J. and J.E. Brazier, *Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D.* Quality of life research, 2005. **14**(6): p. 1523-1532.
11. Harvie, H.S., et al., *Responsiveness and minimally important difference of SF-6D and EQ-5D utility scores for the treatment of pelvic organ prolapse.* Am J Obstet Gynecol, 2019. **220**(3): p. 265.e1-265.e11.
12. Luo, N., J.A. Johnson, and S.J. Coons, *Using instrument-defined health state transitions to estimate minimally important differences for four preference-based health-related quality of life instruments.* Medical care, 2010: p. 365-371.
13. Schünemann, H.J. and G.H. Guyatt, *Commentary—goodbye M (C) ID! Hello MID, where do you come from?* Health services research, 2005. **40**(2): p. 593-597.
14. King, M.T., *A point of minimal important difference (MID): a critique of terminology and methods.* Expert review of pharmacoeconomics & outcomes research, 2011. **11**(2): p. 171-184.
15. Kluzek, S., B. Dean, and K.A. Wartolowska, *Patient-reported outcome measures (PROMs) as proof of treatment efficacy.* BMJ Evidence-Based Medicine, 2021: p. bmjebm-2020-111573.
16. Churruca, K., et al., *Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues.* Health Expect, 2021. **24**(4): p. 1015-1024.
17. Polonsky, W.H., et al., *Assessment of diabetes-related distress.* Diabetes care, 1995. **18**(6): p. 754-760.
18. Johnston, B.C., et al., *Minimally important difference estimates and methods: a protocol.* BMJ open, 2015. **5**(10): p. e007953.
19. Jayadevappa, R., R. Cook, and S. Chhatre, *Minimal important difference to infer changes in health-related quality of life-a systematic review.* J Clin Epidemiol, 2017. **89**: p. 188-198.
20. Coretti, S., M. Ruggeri, and P. McNamee, *The minimum clinically important difference for EQ-5D index: a critical review.* Expert Rev Pharmacoecon Outcomes Res, 2014. **14**(2): p. 221-33.
21. Walters, S.J. and J.E. Brazier, *What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D.* Health and quality of life outcomes, 2003. **1**(1): p. 1-8.

22. Shiroiwa, T., et al., *Japanese population norms for preference-based measures: EQ-5D-3L, EQ-5D-5L, and SF-6D.* Qual Life Res, 2016. **25**(3): p. 707-19.
23. McGlothlin, A.E. and R.J. Lewis, *Minimal clinically important difference: defining what really matters to patients.* Jama, 2014. **312**(13): p. 1342-1343.
24. Guyatt, G.H., et al., *Methods to explain the clinical significance of health status measures.* Mayo Clin Proc, 2002. **77**(4): p. 371-83.
25. Rai, S.K., et al., *Approaches for estimating minimal clinically important differences in systemic lupus erythematosus.* Arthritis research & therapy, 2015. **17**(1): p. 1-8.
26. Revicki, D., et al., *Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes.* Journal of clinical epidemiology, 2008. **61**(2): p. 102-109.
27. Bächinger, D., R. Mlynski, and N.M. Weiss, *Establishing the minimal clinically important difference (MCID) of the Zurich Chronic Middle Ear Inventory (ZCMEI-21) in patients treated for chronic middle ear disease.* European Archives of Oto-Rhino-Laryngology, 2020. **277**(4): p. 1039-1044.
28. Devlin, N., D. Parkin, and B. Janssen, *Methods for analysing and reporting EQ-5D data*. 2020: Springer Nature.
29. Angst, F., A. Aeschlimann, and J. Angst, *The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies.* J Clin Epidemiol, 2017. **82**: p. 128-136.
30. Wyrwich, K.W., et al., *Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life.* Medical care, 1999: p. 469-478.
31. Wyrwich, K.W., W.M. Tierney, and F.D. Wolinsky, *Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life.* Journal of clinical epidemiology, 1999. **52**(9): p. 861-873.
32. Copay, A.G., et al., *Understanding the minimum clinically important difference: a review of concepts and methods.* The Spine Journal, 2007. **7**(5): p. 541-546.
33. Watt, J.A., et al., *Using a distribution-based approach and systematic review methods to derive minimum clinically important differences.* BMC medical research methodology, 2021. **21**(1): p. 41-41.
34. Norman, G.R., J.A. Sloan, and K.W. Wyrwich, *Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation.* Medical care, 2003: p. 582-592.
35. Cohen, J., *Statistical power analysis for the behavioral sciences, Stat.* Power Anal. Behav. Sci, 1988. **567**.
36. Houchen-Wolloff, L. and R.A. Evans, *Unravelling the mystery of the 'minimum important difference' using practical outcome measures in chronic respiratory disease.* Chronic Respiratory Disease, 2019. **16**: p. 1479973118816491.
37. Mouelhi, Y., et al., *How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods.* Health and Quality of Life Outcomes, 2020. **18**(1): p. 1-17.
38. Munn, Z., et al., *Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach.* BMC medical research methodology, 2018. **18**(1): p. 1-7.
39. Mays, N., E. Roberts, and J. Popay, *Synthesising research evidence*, in *Studying the organisation and delivery of health services*. 2004, Routledge. p. 200-232.
40. Colquhoun, H.L., et al., *Scoping reviews: time for clarity in definition, methods, and reporting.* Journal of clinical epidemiology, 2014. **67**(12): p. 1291-1294.
41. Arksey, H. and L. O'Malley, *Scoping studies: towards a methodological framework.* International journal of social research methodology, 2005. **8**(1): p. 19-32.

42. Peters MDJ, G.C., McInerney P, Munn Z, Tricco AC, Khalil H., *Methodology for jbi scoping reviews. In E. Aromataris (Ed.), The Joanna Briggs Institute Reviewers manual 2015 (pp. 3 - 24).* J.B. Institute., Editor. 2015.
43. Böckmann, D., et al., *Cost-Effectiveness of Pulmonary Rehabilitation in Patients With Bronchial Asthma: An Analysis of the EPRA Randomized Controlled Trial.* Value Health, 2021. **24**(9): p. 1254-1262.
44. Brusco, N.K., et al., *Is cost effectiveness sustained after weekend inpatient rehabilitation? 12 month follow up from a randomized controlled trial.* BMC Health Serv Res, 2015. **15**: p. 165.
45. Fueki, K., et al., *Cost-effectiveness analysis of prosthetic treatment with thermoplastic resin removable partial dentures.* J Prosthodont Res, 2021. **65**(1): p. 52-55.
46. Goldstein, R.S., et al., *Economic analysis of respiratory rehabilitation.* Chest, 1997. **112**(2): p. 370-9.
47. McKenna, G., et al., *Cost-effectiveness of tooth replacement strategies for partially dentate elderly: a randomized controlled clinical trial.* Community Dent Oral Epidemiol, 2014. **42**(4): p. 366-74.
48. Verberkt, C.A., et al., *Cost-effectiveness of sustained-release morphine for refractory breathlessness in COPD: A randomized clinical trial.* Respir Med, 2021. **179**: p. 106330.
49. Hays, R.D. and J.M. Woolley, *The concept of clinically meaningful difference in health-related quality-of-life research.* Pharmacoeconomics, 2000. **18**(5): p. 419-423.
50. Briggs, A.H. and B.J. O'Brien, *The death of cost-minimization analysis?* Health Econ, 2001. **10**(2): p. 179-84.
51. Whitehurst, D.G. and S. Bryan, *Trial-based clinical and economic analyses: the unhelpful quest for conformity.* Trials, 2013. **14**(1): p. 1-3.
52. Feng, Y.S., et al., *Psychometric properties of the EQ-5D-5L: a systematic review of the literature.* Qual Life Res, 2021. **30**(3): p. 647-673.
53. Bilbao, A., et al., *Psychometric properties of the EQ-5D-5L in patients with major depression: factor analysis and Rasch analysis.* J Ment Health, 2021: p. 1-11.
54. Yang, Y., J. Brazier, and L. Longworth, *EQ-5D in skin conditions: an assessment of validity and responsiveness.* Eur J Health Econ, 2015. **16**(9): p. 927-39.
55. Marks, M., C. Grobet, and L. Audigé, *Validity, responsiveness and minimal important change of the EQ-5D-5L in patients after rotator cuff repair, shoulder arthroplasty or thumb carpometacarpal arthroplasty.* Qual Life Res, 2021. **30**(10): p. 2973-2982.
56. Olerud, P., et al., *Responsiveness of the EQ-5D in patients with proximal humeral fractures.* J Shoulder Elbow Surg, 2011. **20**(8): p. 1200-6.
57. Globe, G., et al., *Evaluating minimal important differences and responder definitions for the asthma symptom diary in patients with moderate to severe asthma.* Journal of Patient-Reported Outcomes, 2019. **3**(1): p. 1-16.
58. Hays, R.D. and J.D. Peipert, *Minimally important differences do not identify responders to treatment.* JOJ scin, 2018. **1**(1): p. 555552.
59. Mulhern, B., et al., *Comparing the UK EQ-5D-3L and English EQ-5D-5L Value Sets.* PharmacoEconomics, 2018. **36**(6): p. 699-713.