"NOTICE: this is the author's version of a work that was accepted for publication in *Value in Health*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Value in Health*, *Volume 23, Issue 7, July 2020, Pages 936-944*, https://doi.org/10.1016/j.jval.2020.03.003

Also, add this text to the title page of the postprint: © 2020. This manuscript version is made available under the CC-BY-NC-ND 4.0 licence http://creativecommons.org/licenses/by-nc-nd/4.0/ 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

Running Title: Minimally Important Difference in EQ-5D Score [See full title above]

Edward B. Henry, MSc

J.E. Cairnes School of Business and Economics, NUI Galway, Galway, Republic of Ireland Luke E. Barry, MSc

Centre for Public Health, Queen's University of Belfast, Belfast, Northern Ireland

Anna P. Hobbins, PhD

Centre for Research in Medical Devices (Cúram) and Health Economics and Policy Analysis Centre (HEPAC), NUI Galway, Galway, Ireland.

Nathan S. McClure, MSc

School of Public Health, University of Alberta, Edmonton, Alberta, Canada

Ciaran O'Neill, PhD

Centre for Public Health, Queen's University of Belfast, Belfast, Northern Ireland

Corresponding Author: Ciaran O'Neill, PhD

Centre for Public Health, School of Medicine, Dentistry and Biomedical Sciences, Institute of Clinical Sciences, Block B, Queens University Belfast, Royal Victoria Hospital, Belfast, BT12 6BA, United Kingdom, Ciaran.ONeill@qub.ac.uk, Phone: +44 (0)28 9097 8928 Fax: +44 (0) 28 9023 5900

Funding/Support: This study was supported by the Health Research Board under a Research Leader Award held by Ciaran O'Neill RL/2013/16.

Precis: Simulation-based instrument-defined estimation of the minimally important difference in EQ-5D-5L index score for scoring algorithms developed using EQ-VT version 2 protocols.

Acknowledgements: This work was funded under a Health Research Board Research Leader Award RL/2013/16. The authors would also like to acknowledge Dr. Kieran Walsh and Dr. Stephen O'Neill for comments on earlier drafts of this manuscript and the methods therein.

Word Count: 4248

Number of Pages: 28

Number of Figures: 2

Number of Tables: 2

Appendix: 1

Pages: 1

Figures: 0

Tables 1

Abstract

Objectives: Estimate and compare the MID in index score of country-specific EQ-5D-5L scoring algorithms developed using EuroQol Valuation Technology protocol version 2 including algorithms from Germany, Indonesia, Ireland, Malaysia, Poland, Portugal, Taiwan and the United States.

Methods: A simulation-based approach contingent on all single-level transitions defined by the EQ-5D-5L descriptive system was used to estimate the MID for each algorithm.

Results: The resulting mean (and standard deviation, SD) instrument-defined MID estimates were: Germany, 0.083 (0.022); Indonesia, 0.093 (0.012); Ireland, 0.098 (0.023); Malaysia, 0.072 (0.010); Poland, 0.080 (0.030); Portugal, 0.080 (0.018); Taiwan, 0.101 (0.010); and USA, 0.078 (0.014).

Conclusions: These population preference-based MID estimates and accompanying evidence of how such values vary as a function of baseline index score can be used to aid interpretation of index score change. The marked consistency in the relationship between the calculated MID estimate and the range of the EQ-5D-5L index score, represented by a ratio of 1:20, might substantiate a rule of thumb allowing for MID approximation in EQ-5D-5L index score warranting further investigation.

Keywords: EQ-5D-5L, utility values, health state preference, minimally important difference

Highlights

- This paper reports simulation-based instrument-defined estimates of the minimally important difference in EuroQol 5-dimension 5-level questionnaire (EQ-5D-5L) index score for several country-specific scoring algorithms developed using the EQ-VT version 2 valuation protocols.
- These estimates can be used to aid interpretation of the changes in scale score in the context of clinical trials, longitudinal studies, population health monitoring and economic analyses among others.
- The relative uniformity of the relationship between the MID estimate and country tariff range, in the ratio of 1:20, is in contrast to that observed with the 3L and may reflect greater consistency of approach imposed by iterative refinements to the EQ-VT protocol. This congruence might facilitate development of a rule of thumb for MID estimation of other scoring algorithms.

Introduction

The construction of quality-adjusted life-years (QALYs) necessitates an instrument which provides a descriptive system to operationalize the measurement of an individual's health and a means of assigning weights to this description in a manner reflective of utility values [1]. The EuroQol 5-dimension questionnaire offers one such method of operationalization [2]. It is the preferred generic instrument in the UK among many other countries and is used to inform decisions on the allocation of large amounts of healthcare resources [3-5]. While there are concerns over the elicitation methods employed in developing value sets for the EQ-5D-5L, iterative refinements have resulted in greater confidence in tariffs produced using more recent version 2.0 and 2.1 protocols of the standardised EuroQol Valuation Technology (EQ-VT). These give a firmer basis for cross-national comparison, with variations in reported preferences between countries more likely to reflect preference divergence rather than differences pertaining to study design [6].

The over-reliance on the statistical significance of findings, seen as problematic in general within the scientific community, is all the more pernicious in relation to such health-related quality of life (HRQoL) instruments given their multidimensional nature [7, 8]. This concern combined with the prevalent use of EQ-5D instruments has highlighted the need to establish what constitutes a meaningful change in score of such measures [8, 9]. Moreover, while experience and familiarity provide clinicians and researchers with the means to make purposeful inference on the basis of physiological measures, generic HRQoL measures, such as EQ-5D index scores, may be less familiar to those tasked with their interpretation [10]. Accordingly, there have been increased efforts to establish the minimally important difference (MID) in HRQoL scores [9, 11-15]. In this study, the MID is defined as the smallest difference in EQ-5D-5L index score, based on country tariffs developed using the EQ-VT version 2 protocols, considered an unambiguous change either for the better or worse [9, 15].

The practical uses for such MID estimates are diverse. In the context of individual patients, MIDs may aid the determination of the amount of change in a HRQoL measure which is sufficient to justify treatment change, be that commencement of a new treatment, continuing or stopping of existing treatments or changing treatment dose/intensity [8]. Third party payers may also seek to employ MID estimates in decisions related to continuation or cessation of reimbursement for care or to audit the performance of service providers [13]. Moreover, for clinical researchers, MIDs can be used to establish *a priori* power, sample size determination and the ascertainment of a therapeutic threshold signalling an intervention's effectiveness [8, 12, 13].

There are two primary means by which one might estimate the MID: distribution-based and anchor-based approaches. The former makes use of distributional properties of the outcome score (e.g., one-half standard deviation) as a method to approximate the MID [8, 16, 17]. The main criticism of this approach is that it does not directly consider whether the observed change is meaningful or minimally important from the perspective of patients [9, 15, 18]. By contrast, the anchor-based approach involves the appraisal of the change in the HRQoL score in relation to an associated external anchor or criterion such as a clinical endpoint or global rating of change scale [19]. Anchor-based methods may be limited by several practical challenges related to the choice of anchor, consensus on what constitutes the smallest important change and the resource intensive/time consuming nature of collection of primary data [15]. Owing to their ability to account for patient and/or clinician perspectives, anchor-based methods using multiple anchors are generally recommended, with distribution-based estimates used to support and inform interpretation where appropriate [20].

A novel alternative comprising a simulation-based approach wherein single-level transitions described by the instrument represent minimally important changes is beneficial since these transitions constitute multiple internal anchors obviating immediate requirements for external

anchors [15]. Though this approach has been employed to estimate EQ-5D-3L and EQ-5D-5L index score MIDs for some countries, with demonstrable validity, to our knowledge there are no published estimates for scoring algorithms developed using EQ-VT version 2 protocols [9, 15]. These MID estimates with accompanying evidence of how such values might vary as a function of baseline index score may be used to aid interpretation of index score change prior to context-specific estimation efforts. Accordingly, this study aims to generate simulation-based estimates of the MID in EQ-5D-5L for a range of countries' value sets produced using the EQ-VT v2.0 or 2.1 including Germany, Indonesia, Ireland, Malaysia, Poland, Portugal, Taiwan and the United States.

Methods

Included country-specific value sets

EQ-5D-5L valuation studies conducted using EQ-VT version 2 protocols were reviewed and coefficients of preferred models extracted. Index scores for health states are calculated by subtracting the sum of the applicable coefficients (i.e. disutility associated with that health state), as well as the constant term where applicable, from one.

A hybrid model (cTTO and DCE) was the preferred model for all countries except for the US value set which was TTO based [1, 21-27]. Summary statistics of the individual country-specific value sets are presented in Table 1.

INSERT TABLE 1 HERE

Instrument-defined simulation-based MID estimation

Estimation using an instrument-defined approach enumerates the smallest health transitions described by the EQ-5D-5L, namely, single-level transitions with the difference in index score of the two considered health states representing an individual MID value for a single (internal) anchor. These single-level transitions are defined as a change from the baseline state to a directly adjacent level of a single dimension [9, 15]. Figure 1 exemplifies all possible single-level transitions from a baseline state.

INSERT FIGURE 1 HERE

For each of the 3125 baseline states and accompanying transition states, index scores are calculable using the respective country scoring algorithm allowing for computation of the absolute difference between baseline and transition state index scores. These differences are averaged yielding the mean transition value or single-state MID estimate. Using the health state 44444, for example, the index score calculated using the Irish scoring algorithm is -0.506. Single-level transitions from this state to directly adjacent health states 34444, 43444, 44344,

44434, 44443, 54444, 45444, 44544, 44454, 44445 yield absolute differences in index score of 0.118, 0.141, 0.082, 0.28, 0.333, 0.129, 0.058, 0.033, 0.137 and 0.111 respectively. This results in a mean transition value of 0.142. This process is repeated by simulation for all health states and these mean transition values are again averaged to establish a point estimate of the instrument-defined MID. The EQ-5D-5L instrument allows for 25,000 transitions in total [15, 28].

Implicit in this procedure is the assumption that any level change constitutes an important difference in health with single-level transition representing an MID [9]. However, some single-level transitions may invoke an index score change considered atypical of the magnitude of change in an MID value, and could be, therefore, excluded from estimation of the MID [9, 15]. These "maximum-valued scoring parameters" are defined as the largest shift in index score resulting from a single-level transition within each dimension. Examining the Irish scoring algorithm, a transition between level 3 and 4 in the dimension of pain/discomfort results in an absolute difference in index score of 0.280 which represents one of the maximum-valued scoring parameters excluded. This exclusion criterion is repeated for all dimensions of the instrument, excluding transitions resulting in the 5 maximum-valued scoring parameters, 1 from each dimension. In the case of the German scoring algorithm, for example, exclusion of single-level transitions invoking maximum-valued scoring parameters (i.e. those transitions between levels 3 and 4 in any dimension) allows for only 5 possible single-level transitions from a baseline state of 44444, to adjacent states 54444, 45444, 44544, 44454 and 44445. Repeating this exclusion for all applicable baseline states reduces the feasible transition total to 18,750. From this total the MID estimate is aggregated [9, 15].

Similarly, certain minimally-valued scoring parameters describing the smallest possible absolute difference in index score resulting from a single-level transition within a dimension, may be less than minimally important [9, 15]. Using the example of the pain/discomfort

dimension of the Indonesian value set, the smallest absolute difference in index score is 0.009 which results from a transition between level 2 and 3. Exclusion of the minimum-valued scoring parameters from each dimension of the instrument also results in 18,750 possible transitions. The maximum and minimum-valued scoring parameters for each country's value set are displayed in Table 1.

For illustrative purposes, the parameter-excluded MID estimates resulting from these sensitivity analyses are presented with the MID estimate based on all transitions garnering one feasible interval within which the MID for the EQ-5D-5L might fall. An alternative approach, construction of a 95% confidence interval based on MID estimates calculated following 1000 random draws from the probability distributions of the utility values from the respective country tariffs, was also undertaken to account for the uncertainty in the increments or decrements to utility associated with those point estimates.

Scatterplots showing the distribution of mean transition values as a function of baseline index score were produced. Locally weighted regression (LOESS) was employed as an abbreviated graphical summary of the relationship between the dependent variable, the individual health state MID estimate, and independent variable, the baseline EQ-5D-5L index score of that health state. This method is non-parametric, the fitted curve being obtained empirically rather than through pre-specification of the functional form of the relationship of the variables [29]. A proportion of the total dataset, in this case 80%, is included in a series of weighted regressions. Although the choice of proportion is, largely, discretionary the general objective is to produce a smooth curve which retains the structure inherent within the data [29]. The observations are inversely weighted according to distance from the evaluation point on the *x*-axis using the tricube function [30]. Weighting ensures that observations closer to the evaluation point have a greater influence on the placement of the predicted value than those falling further away within the local region [29]. The fitted values, based on the coefficients resulting from each

local regression, are superimposed over the scatterplot and connected by line segments [15, 29]. The LOESS curves and scatterplot distributions allow for visual appraisal of the consistency of calculated mean transition value across health state severity [28].

To summarise the relationship between MID estimate and corresponding tariff range (sometimes referred to as the scale length), their ratios were computed. For each scoring algorithm this range was obtained by calculating the difference in index score between the worst (55555) and best (11111) possible health states. Instrument-defined MID estimates were also quantified for a range of EQ-5D-3L scoring algorithms and they are presented in the appendix of this article together with the ratio of that tariff range to the MID estimate. Also presented in this appendix are the ratios of previously published instrument-defined MID estimates. All analyses were conducted using Microsoft Excel (2016) and STATA/SE 15.1 (Stata Corp, College Station, TX, USA).

Results

Germany

Many transitions between states result in the same value of index score change. In the case of Germany, after removing these duplicates, 1685 unique mean health state transition values were calculable, ranging in absolute value from 0.029 to 0.148. The aggregated mean (standard deviation, SD) MID estimate was 0.083 (0.022) (Table 2). The distribution of the scatterplot and slope of the LOESS curve suggest a negative correlation between these estimates and baseline index score (Figure 2). As the baseline gets lower, the MID gets larger.

Maximum and minimum-valued parameter exclusion resulted in shifts within the distribution providing an interval of 0.062 to 0.101 for the point MID estimate as shown in Table 2. While the related LOESS curves are still suggestive of a negative correlation, exclusion of minimum-valued parameters appears to introduce some uniformity to MID estimates based on baseline health states with an index score of -0.20 or less (Figure 2).

Indonesia

This tariff yields 1245 individual MID estimates which range from 0.071 to 0.131. In this case, the summarised mean (SD) MID estimate was 0.093 (0.012) (Table 2). The MID estimate remains constant below an index score of -0.20 with some indication of smaller MIDs relating to health states with larger index scores thereafter (Figure 2).

The minimum and maximum-excluded MID estimate-based interval was 0.080 to 0.108 with distributional displacements concentrated towards the centre of the scatterplot (Table 2). Excluding minimum-valued scoring parameters results in a LOESS curve indicative of slightly larger MID estimates for health states with mid-range index scores. The opposite can be

inferred from the distribution of MID estimates calculated after exclusion of maximum-valued scoring parameters (Figure 2).

INSERT TABLE 2 HERE

Ireland

1767 distinct single-state MID estimates were identifiable, ranging from 0.048 to 0.158. The summarised mean (SD) of these MID estimates was 0.098 (0.023) (Table 2). The plotted distribution and LOESS suggest that smaller MID estimates are calculable from health states with higher index scores (Figure 2).

Exclusion of transitions with maximum and minimum-valued scoring parameters provided an interval of 0.068 to 0.116 (Table 2). The distribution resulting from the maximum-valued scoring parameter exclusion suggests that omitting these transitions tempered the relationship between index score and MID estimate. MID values calculated after exclusion of minimum-valued scoring parameters appear consistent across baseline index scores up to 0.30 approximately, after which, the MID estimate is smaller for health states with higher index scores (Figure 2).

Malaysia

This tariff produced 1118 individual health state MID estimates ranging from 0.046 to 0.099 and resulted in an aggregated mean (SD) MID estimate of 0.072 (0.010) (Table 2). Examining the scatterplot distribution and LOESS, the relatively narrow variation in MID estimates across all baseline index scores is indicative of a slight decrease in predicted MIDs at higher baseline index scores (Figure 2).

An interval of 0.053 to 0.088 resulted from MID re-estimation following parameter exclusion (Table 2). The minimum-valued scoring parameter exclusion resulted in a distributional shift

to larger MID estimates with estimates at the upper and lower limits of baseline index score slightly below mid-range estimates. Exclusion of maximum-valued scoring parameter transitions resulted in MID estimates at the index score extremes which were marginally higher than those at the centre (Figure 2).

INSERT FIGURE 2 HERE

Poland

1982 distinct mean transition values ranging from 0.02 to 0.183 were identifiable. The summarised mean (SD) was 0.080 (0.030) (Table 2). The scatterplot and LOESS curve reveal a negative correlation between the MID estimate and index score (Figure 2).

Maximum and minimum-valued parameter exclusion led to aggregated MID estimates of 0.048 and 0.099 respectively (Table 2). There was only a minor distributional shift resulting from the minimum-valued scoring parameter exclusion, the LOESS curve consistent in shape with the curve observed based on all transitions. By contrast, the maximum-valued scoring parameter exclusion curbed the negative association (Figure 2).

Portugal

This tariff resulted in 1635 unique single-state MID estimates which ranged from 0.037 to 0.136. The aggregated mean (SD) MID estimate was 0.080 (0.018) (Table 2). The scatterplot is reminiscent of Polish and German scatterplots with higher MID estimates possible for baseline states with lower index scores (Figure 2).

Excluding transitions with maximum and minimum-valued scoring parameters produced an interval of 0.060 to 0.095 (Table 2). The minimum-valued scoring parameter exclusion produced an upward shift in plot distribution, still indicative of the aforementioned negative correlation. While the moderating effect of exclusion of transitions with maximum-valued

scoring parameters appears to be smaller than that observed in relation of the Polish tariff, an attenuated association is still perceivable (Figure 2).

Taiwan

The scoring algorithm for Taiwan produced 2521 unique MID estimates ranging from 0.073 to 0.128. The summarised mean (SD) was 0.101 (0.010) (Table 2). The plot of MID estimates by index score produced a LOESS curve with a slight negative gradient approximating around the mean MID estimate (Figure 2).

The summarised mean MID estimates produced after parameter exclusion were 0.082 and 0.112 respectively (Table 2). Both types of exclusion resulted in LOESS curves suggestive of an invariability in MID estimates across most baseline index scores with some evidence of a small reduction in minimally-excluded MID estimates relating to baseline index scores above 0.40 (Figure 2).

United States

In this case there were 1380 distinct MID estimates ranging from 0.040 to 0.116. The aggregated mean (SD) was 0.078 (0.014) (Table 2). The scatterplot and LOESS curve revealed that the MID estimate is comparatively stable across all baseline index scores, adhering to the summarised mean MID (Figure 2). MID estimates calculated after exclusion of transitions with maximum and minimum-valued scoring parameters resulted in an interval of 0.053-0.098 (Table 2). LOESS curves reveal uniformity in MID estimates for negative baseline index scores. Thereafter, the curves converge on a similar MID estimated on the basis of all transitions at a baseline index score of 1.00 (Figure 2).

Also illustrated in Table 2, the MID estimate (based on all transitions) to index score range ratio was 1:20 for all included country tariffs.

Discussion

This analysis implements the instrument-defined approach first described by Luo *et al.* [9] to estimate the MID in EQ-5D-5L index score for scoring algorithms developed from country-specific utility values elicited using EQ-VT version 2 protocols. The MID estimates ranged from 0.072 for the Malaysian tariff to 0.101 in the case of Taiwan with German, Indonesian, Irish, Polish, Portuguese and US values falling between these bounds. These estimates are slightly larger in magnitude than EQ-5D-5L MID values from previous instrument-defined estimation studies which ranged from 0.048 to 0.069 (Appendix 1) with differences reflective of the larger index score ranges of the value sets included in this study.

Our findings are consistent with other published studies estimating the EQ-5D-5L MID. In Taiwanese patients undergoing stroke rehabilitation, Chen *et al.* [31] employed an anchorbased approach where the change in EQ-5D-5L index score corresponding to a 10-15% change in perceived recovery score of the Stroke Impact Scale 3.0 constituted an MID. Despite limitations related to the use of index scores based on a crosswalk of a Japanese EQ-5D-3L value set and estimation based on a single external anchor, the reported MID of 0.10 is consistent with the Taiwanese instrument-defined MID estimate of 0.101 contained herein. As part of a comparison of US EQ-5D-3L and EQ-5D-5L value sets, investigators computed an MID estimate of 0.07 in index score from model coefficients established via a main effects specification [28]. Our estimated MID of 0.078 calculated using coefficients from the preferred cTTO tobit model bore marked similarity to their findings. This confluence of estimates demonstrates their robustness and adds weight to calls for empirically derived MID values to be evaluated in the context of instrument-defined MIDs [9, 15].

Motivated by the widespread use of HRQoL instruments, the challenge of interpretation and concern about over-reliance on statistical significance, a range of approaches have been put forward to establish what might be the smallest change in the score of such measures considered meaningful. Although there is still no consensus on a "gold standard" of MID estimation it is generally recommended that anchor-based estimates be assigned most weight with triangulation of findings from multiple anchors and other approaches where possible [20]. Its capacity to provide a summarised estimate of multiple internal anchors based on population preferences means that the instrument-defined approach could be viewed as complying with these recommendations. Instrument-defined estimates can therefore aid interpretation of index score change in settings such as economic analyses, clinical trials, cohort studies and population health monitoring [14]. The interpretation of EQ-5D-5L index score change has already been informed by instrument-defined estimates in randomised controlled trials examining treatments for ovarian cancer, hypothyroidism and multiple myeloma [32-34]. That said, factoring instrument-defined estimates into a wider consideration of estimates garnered from other approaches is advisable, as per the triangulation of multiple estimates recommended by Revicki *et al.* [20]. As with the development of the EQ-5D-5L instrument itself, determination of associated MID estimates could be described as an iterative process with new evidence building on approaches and estimates already described in the literature.

The aggregated mean MID estimates enumerated here are more fully supported by a stable or flat distribution of single-state mean transition values which were observed in relation to certain country tariffs, as evidenced by flat LOESS curves in Figure 2. Non-linearity in utility values with respect to severity may result in less stability in the point estimate. For example, the distribution of single-state MID estimates pertaining to the Portuguese tariff indicates that if the patient group under consideration has a baseline index score of 0.20 an MID of approximately 0.10 is appropriate, whereas a baseline score of 1 would entail the use of an MID of 0.04. In these scenarios, a more nuanced interpretation may be gained by consideration of the range of parameter-excluded estimates which omit the greatest of these utility discontinuities in calculation of the overall average. In sensitivity analysis here, exclusion of

transitions with maximum-valued scoring parameters generated lower aggregated mean MID estimates consistent with previously published results [15]. Alternatively, examination of how the point estimate varies as a function of baseline index score, provided by the accompanying LOESS curve, may give the user greater insight as to the appropriateness of the overall MID estimate to their specific scenario.

Evidently, an awareness of context is key in evaluation of whether an observed change is meaningful. Empirical appraisal of context-specific transition likelihoods and the use of these transition probabilities to weight individual health state MID estimates is another potential avenue for future research. An estimation of the EQ-5D-5L index score MID in adults with type 2 diabetes has been undertaken by McClure *et al.* [35] but, naturally, there are many disease groups which warrant investigation. Considering that countries such as the Netherlands, UK and Ireland support the use of EQ-5D tools in economic evaluations [4, 5, 36], further research may also examine how certain interventions differ in their ability to produce an index score change deemed to be minimally important.

The EQ-5D-3L instrument-defined MID estimates appear higher than those from EQ-5D-5L scoring algorithms. Their application would mean that any observed index score change needs to be relatively large before being considered minimally important. This observation fits with the notion, expressed in literature, of the EQ-5D-3L instrument being less discriminatory [1, 37-39] and highlights the value of sensitivity analysis which excludes transitions with maximum-valued scoring parameters [9].

There appears to be a marked consistency in the relationship between the approximated MID value and the range of the index score. This ratio of 1:20 might allow for formulation of a rule of thumb where division of the index score range by 20 produces an estimate of the EQ-5D-5L MID. This uniformity is not found in the corresponding ratios of EQ-5D-3L instrument-defined MID estimate to tariff range presented in the appendix. A plausible explanation for this is that

refinements to EQ-5D-5L valuation protocols provide greater response consistency across individuals through, for example, the inclusion of practice states, quality control procedures to ensure compliance and structured feedback [6]. That the valuation technology may be responsible for the observed congruity is also supported by the fact the ratio persists for the US EQ-5D-5L MID which used an alternative modelling strategy. It follows that one might express a greater degree of confidence in a rule of thumb proposed on the basis of scoring algorithms developed using EQ-VT version 2 protocols. Nonetheless, application of any rule of thumb necessitates due caution and further investigation is required.

A number of caveats to the use of instrument-defined MID estimates should be borne in mind. Firstly, there is a tendency for MIDs to vary by disease group/severity, patient baseline status, direction of change and demographic factors [13, 15, 20]. Compounding this issue is the psychological adaption to illness and resulting "response shift" in HRQoL which may also differ by study population [8]. It must also be noted that the procedure cannot be applied to psychometrically scored HRQoL instruments which are primarily scored using multiple items equally weighted.

Although it is possible to speculate as to the factors underlying between-country differences in MID estimates, differentiation of the contribution of each potential source of variation was beyond the scope of this investigation. As such, cultural and demographic distinctions reflective of true variation in health preferences and attitudes to death are confounded by differences in modelling strategy, sampling discrepancies and reported violations of EQ-VT version 2 protocols [1, 15, 21-27, 40]. Furthermore, some distributional patterns and LOESS curve shapes relating the individual single-state MID estimates to baseline index score can be explained fully by the possible transitions of the instrument and methods comprising the instrument-defined approach. Where intermediate index scores yield higher MID estimates than at the extremes of the range, McClure *et al.* [35] posit that this is simply a reflection of

extremity-based MID estimates being more representative of transitions in a single direction. Consequently, MID estimates from the centre of the range are reflective of a more varied assortment of transitions leading to larger MID estimates. In other instances, identifying the causes of these variations becomes more complex with the shape of the distribution likely due, in part, to the factors which also underpin the distinctions in country-specific MID estimates. It is also worth mentioning that, while we employ terminology consistent with previous literature regarding important changes in health measures for clinical decision making, the term 'minimally important difference' could be misleading, at least with regard to instrumentdefined approach used here and elsewhere. Given the nature of the calculations inherent in the approach, a more accurate reflection may be an 'average important difference'.

Conclusions

This study used the simulation-based instrument-defined approach to estimate the MID in EQ-5D-5L index score for scoring algorithms based on the utility values of Germany, Indonesia, Ireland, Malaysia, Poland, Portugal, Taiwan and the United States. These methods can be extended to other value sets as they become available. The resulting estimates may inform interpretation of index score change by clinicians and researchers alike, in trial design, economic evaluations and decisions pertaining to intervention effectiveness at the individual patient level. That said, caution is warranted, given the value set-specific propensity of the MID estimate to fluctuate with baseline index score.

References

- Hobbins, A., et al., Utility Values for Health States in Ireland: A Value Set for the EQ-5D-5L. Pharmacoeconomics, 2018. 36(11): p. 1345-1353.
- Rabin, R. and F. de Charro, *EQ-5D: a measure of health status from the EuroQol Group*. Ann Med, 2001. **33**(5): p. 337.
- Dakin, H., et al., *The influence of cost-effectiveness and other factors on NICE decisions*. Health Econ, 2015. 24(10): p. 1256-1271.
- NICE. Guide to the methods of technology appraisal 2013. [cited 2018 10 November]; Available from: https://www.nice.org.uk/process/pmg9/resources/guideto-the-methods-of-technology-appraisal-2013-pdf-2007975843781.
- 5. HIQA. Draft Guidelines for the Economic Evaluation of Health Technologies in Ireland. 2017 [cited 2018 19 November]; Available from: https://www.hiqa.ie/sites/default/files/2017 10/Draft_HIQA_Economic_Guidelines_2017_consultation.pdf.
- Stolk, E., et al., Overview, update, and lessons learned from the international EQ-5D-5L valuation work: Version 2 of the EQ-5D-5L Valuation Protocol. Value Health, 2019. 22(1): p. 23-30.
- Amrhein, V., S. Greenland, and B. McShane, *Retire statistical significance*. Nature, 2019. 567.
- 8. King, M.T., *A point of minimal important difference (MID): a critique of terminology and methods.* Expert Rev Pharmacoecon Outcomes Res, 2011. **11**(2): p. 171-184.
- 9. Luo, N., J. 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. Med Care, 2010. **48**(4): p. 365-71.

- 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*. Qual Life Res, 2005. 14(6): p. 1523-32.
- 11. Crosby, R.D., R.L. Kolotkin, and G.R. Williams, *Defining clinically meaningful change in health-related quality of life*. J Clin Epidemiol, 2003. **56**(5): p. 395-407.
- 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. 221-233.
- Wright, A., et al., *Clinimetrics corner: a closer look at the minimal clinically important difference (MCID)*. J Man Manip Ther, 2012. 20(3): p. 160-166.
- 14. ISPOR. Minimal Clinically Important Difference in EQ-5D: We Can Calculate it But Does That Mean We Should? 2017 [cited 2019 27 May]; Available from: https://www.ispor.org/docs/default-source/presentations/1066.pdf?sfvrsn=25feffd6_1.
- McClure, N.S., et al., Instrument-Defined Estimates of the Minimally Important Difference for EQ-5D-5L Index Scores. Value Health, 2017. 20(4): p. 644-650.
- 16. Norman, G.R., J.A. Sloan, and K.W. Wyrwich, *Interpretation of changes in healthrelated quality of life: the remarkable universality of half a standard deviation.* Med Care, 2003: p. 582-592.
- Hays, R.D., S.S. Farivar, and H. Liu, *Approaches and recommendations for estimating minimally important differences for health-related quality of life measures*. COPD, 2005. 2(1): p. 63-67.
- Dvir, Z., Difference, significant difference and clinically meaningful difference: The meaning of change in rehabilitation. J Exerc Rehabil, 2015. 11(2): p. 67.

- 19. Jaeschke, R., J. Singer, and G.H. Guyatt, *Measurement of health status: ascertaining the minimal clinically important difference*. Control Clin Trials, 1989. 10(4): p. 407-415.
- Revicki, D., et al., *Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes*. J Clin Epidemiol, 2008. 61(2): p. 102-109.
- 21. Ludwig, K., J.-M.G. von der Schulenburg, and W. Greiner, *German value set for the EQ-5D-5L*. Pharmacoeconomics, 2018: p. 1-12.
- 22. Purba, F.D., et al., *The Indonesian EQ-5D-5L value set*. Pharmacoeconomics, 2017. **35**(11): p. 1153-1165.
- Shafie, A.A., et al., *EQ-5D-5L Valuation for the Malaysian Population*.
 Pharmacoeconomics, 2018: p. 1-11.
- 24. Golicki, D., et al., *Valuation of EQ-5D-5L Health States in Poland: the First EQ-VT-Based Study in Central and Eastern Europe.* Pharmacoeconomics, 2019: p. 1-12.
- 25. Ferreira, P.L., et al., *A hybrid modelling approach for eliciting health state preferences: the Portuguese EQ-5D-5L value set.* Qual Life Res, 2019: p. 1-13.
- Lin, H.-W., et al., *Valuation of the EQ-5D-5L in Taiwan*. PLoS One, 2018. 13(12): p. e0209344.
- Pickard, A.S., et al., United States Valuation of EQ-5D-5L Health States Using an International Protocol. Value Health, 2019.
- Law, E.H., et al., Parallel Valuation: A Direct Comparison of EQ-5D-3L and EQ-5D-5L Societal Value Sets. Med Decis Making, 2018. 38(8): p. 968-982.
- 29. Jacoby, W.G., *Loess: a nonparametric, graphical tool for depicting relationships between variables.* Elect Stud, 2000. **19**(4): p. 577-613.

- Cleveland, W.S., *Robust locally weighted regression and smoothing scatterplots*. J Am Stat Assoc, 1979. 74(368): p. 829-836.
- 31. Chen, P., et al., Validity, responsiveness, and minimal clinically important difference of EQ-5D-5L in stroke patients undergoing rehabilitation. Qual Life Res, 2016.
 25(6): p. 1585-96.
- 32. Moreau, P., et al., *Convenience, satisfaction, health-related quality of life of onceweekly 70 mg/m 2 vs. twice-weekly 27 mg/m 2 carfilzomib (randomized ARROW study).* Leukemia, 2019: p. 1.
- 33. Oza, A.M., et al., Quality of life in patients with recurrent ovarian cancer treated with niraparib versus placebo (ENGOT-OV16/NOVA): results from a double-blind, phase 3, randomised controlled trial. Lancet Oncol, 2018. 19(8): p. 1117-1125.
- Stott, D.J., et al., *Thyroid hormone therapy for older adults with subclinical hypothyroidism*. N Engl J Med, 2017. **376**(26): p. 2534-2544.
- 35. McClure, N.S., et al., *Minimally Important Difference of the EQ-5D-5L Index Score in Adults with Type 2 Diabetes.* Value Health, 2018. **21**(9): p. 1090-1097.
- 36. Versteegh, M., S. Knies, and W. Brouwer, *From good to better: new Dutch guidelines for economic evaluations in healthcare*. 2016, Springer.
- 37. Mulhern, B., et al., *Comparing the UK EQ-5D-3L and English EQ-5D-5L value sets*.
 Pharmacoeconomics, 2018. 36(6): p. 699-713.
- 38. Longworth, L., et al., *Use of generic and condition-specific measures of healthrelated quality of life in NICE decision-making: a systematic review, statistical modelling and survey.* Health Technol Assess, 2014.
- 39. Brazier, J., et al., A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health

populations and the estimation of mapping functions from widely used specific measures. Health Technol Assess, 2014. **18**(34): p. vii.

40. Gerlinger, C., et al., *Comparing the EQ-5D-5L utility index based on value sets of different countries: impact on the interpretation of clinical study results.* BMC Res Notes, 2019. **12**(1): p. 18.

Country	Modelling	Index	Maximum-valued	Minimum-valued scoring parameters	
(EQ-VT		Score	scoring parameters		
protocol)	Арргоасп	Range	(Level transition)	(Level transition)	
			MO – 0.097 (3↔4)	MO – 0.016 (2↔3)	
Germany (v2.0)	Hybrid	-0.661 to 1	SC – 0.113 (3↔4)	$SC - 0.006 (2 \leftrightarrow 3)$	
	(cTTO +		UA – 0.080 (3↔4)	UA – 0.013 (2↔3)	
	DCE)		PD – 0.295 (3↔4)	$PD - 0.052 (2 \leftrightarrow 3)$	
			AD – 0.162 (3↔4)	$AD - 0.030 (1 \leftrightarrow 2)$	
		-0.865 to 1	MO – 0.218 (3↔4)	MO – 0.073 (2↔3)	
Indonesia	Hybrid		SC – 0.108 (3↔4)	$SC - 0.039 (2 \leftrightarrow 3)$	
(v2.0)	(cTTO +		UA – 0.145 (3↔4)	UA – 0.066 (2↔3)	
	DCE)		PD – 0.103 (3↔4)	$PD - 0.009 (2 \leftrightarrow 3)$	
			AD – 0.093 (3↔4)	AD – 0.055 (2↔3)	
Ireland (v2.0)	Hybrid (cTTO + DCE)	-0.974 to 1	MO – 0.129 (4↔5)	MO – 0.034 (2↔3)	
			SC – 0.141 (3↔4)	$SC - 0.033 (2 \leftrightarrow 3)$	
			UA – 0.082 (3⇔4)	UA – 0.023 (2↔3)	
			PD – 0.280 (3↔4)	PD – 0.025 (2↔3)	
			AD – 0.333 (3↔4)	$AD - 0.080 (1 \leftrightarrow 2)$	

Table 1 - Summary statistics of EQ-5D-5L country value sets including dimension-specific maximum and minimum-scoring parameters [1, 21-27]

			MO = 0.153 (3 - 4)	$MO = 0.027(2 \rightarrow 3)$
			$MO = 0.133 (3 \leftrightarrow 4)$	$\mathbf{WIO} = 0.027 \ (2 \leftrightarrow 3)$
Malaysia	Hybrid (cTTO +	-0.442 to 1	SC – 0.117 (3↔4)	SC – 0.021 (2↔3)
			UA – 0.091 (3↔4)	UA – 0.016 (2↔3)
(v2.0)	DCE)		PD – 0.152 (3↔4)	$PD - 0.026 (2 \leftrightarrow 3)$
			AD – 0.135 (3↔4)	AD – 0.023 (2↔3)
			MO – 0.188 (4↔5)	MO – 0.009 (2↔3)
D 1 1	Hybrid	-0.590	SC – 0.153 (4↔5)	$SC - 0.016 (2 \leftrightarrow 3)$
	(cTTO +		UA – 0.108 (4⇔5)	UA – 0.017 (2↔3)
(v2.0)	DCE)	to 1	PD – 0.314 (4↔5)	PD – 0.020 (2↔3)
			AD – 0.124 (4⇔5)	AD – 0.011 (2↔3)
			MO – 0.174 (4↔5)	MO – 0.044 (2↔3)
	Hybrid	-0.603 to 1	SC – 0.138 (4↔5)	$SC - 0.022 (2 \leftrightarrow 3)$
Portugal	(cTTO +		UA – 0.128 (4⇔5)	UA – 0.019 (2↔3)
(v2.0)	DCE)		PD – 0.153 (3↔4)	PD − 0.041 (1 \leftrightarrow 2)
			AD – 0.127 (3↔4)	AD – 0.036 (1↔2)
	Hybrid	-1.026	MO – 0.1656 (3↔4)	MO – 0.092 (2↔3)
Taiwan (v2.0)			SC – 0.1322 (3↔4)	SC – 0.0565 (2↔3)
	(cTTO +		UA – 0.1568 (3↔4)	UA – 0.0508 (2↔3)
	DCE)	to 1	PD – 0.1824 (3↔4)	PD – 0.071 (2↔3)
			AD – 0.1572 (3↔4)	AD – 0.0637 (1↔2)

Table 1 Continued

			MO – 0.115 (3↔4)	MO – 0.026 (2↔3)
	A cTTO	0 572	SC – 0.113 (3↔4)	SC – 0.018 (2↔3)
USA		-0.575	UA – 0.154 (3↔4)	UA – 0.000 (4⇔5)
(v2.0)	tobit	to 1	PD – 0.22 (3↔4)	PD – 0.038 (2↔3)
			AD – 0.176 (3↔4)	AD – 0.022 (4↔5)

Table 1 Continued

Notes: Index score range is from the index score for worst possible health state (55555) to best possible health state (11111) based on each scoring algorithm. cTTO indicates composite time trade-off; DCE, discrete-choice experiment; EQ-5D-5L, EuroQol 5-dimension 5-level instrument; EQ-VT, EuroQoL Valuation Technology; MO, mobility; SC, self-care; UA, usual activities; PD, pain/discomfort; AD, anxiety/depression. Table 2 - Instrument-defined MID estimates and accompanying intervals for EQ-5D-5L country scoring algorithms developed using EQ-VT v2.0 with the ratio of these estimates to EQ-5D-5L country-specific scoring algorithm range.

Country	Mean MID estimate (SD)	Max & min exclusion- derived interval	95% CI	Range of value set	MID divided by range	Max- excluded MID divided by range	Min- excluded MID divided by range
Germany	0.083 (0.022)	0.062-0.101	0.081-0.085	1.661	0.05	0.04	0.06
Indonesia	0.093 (0.012)	0.080-0.108	0.091-0.097	1.865	0.05	0.04	0.06
Ireland	0.098 (0.023)	0.068-0.116	0.096-0.101	1.974	0.05	0.03	0.06
Malaysia	0.072 (0.010)	0.053-0.088	0.072-0.072	1.441	0.05	0.04	0.06
Poland	0.080 (0.030)	0.048-0.099	0.077-0.083	1.59	0.05	0.03	0.06
Portugal	0.080 (0.018)	0.060-0.095	0.079-0.082	1.603	0.05	0.04	0.06
Taiwan	0.101 (0.010)	0.082-0.112	0.098-0.104	2.026	0.05	0.04	0.06
USA	0.078 (0.014)	0.053-0.098	0.076-0.084	1.573	0.05	0.03	0.06

Notes: Range of value set based on published range where available or the difference in EQ-5D-5L index score for the worst possible health state (55555) and the best possible health state (11111). CI indicates confidence interval based on MID estimates developed from 1000 random picks from the probability distributions of the utility values of the respective country tariff; EQ-5D-5L, EuroQol five-dimension five-level instrument; MID, minimally important difference; Max-excluded MID, minimally important difference estimate calculated after exclusion of maximum-valued scoring; Min-excluded MID, minimally important difference estimate calculated after exclusion of minimum-valued scoring; SD, standard deviation.



Figure 1 - All possible single-level transitions from an example EQ-5D-5L baseline health state (44444). 44444 represents a level 4 response in each dimension: mobility, self-care, usual activities, pain/discomfort and anxiety/depression.



Figure 2 - Country-specific scatterplots of individual health state MID estimate as a function of corresponding baseline EQ-5D-5L index score. Scale varies according to the range of index score and/or estimated MIDs. Purple circles and solid-line LOESS curves represent estimates derived from all transitions. Green triangles and dashed-line LOESS curves represent estimates calculated after exclusion of transitions with minimum-valued scoring parameter. Gold crosses and dotted-line LOESS curves represent estimates calculated after exclusion of transitions with minimum-valued after exclusion of transitions with maximum-valued scoring parameters. EQ-5D-5L indicates EuroQol five-dimension five-level instrument; MID, minimally important difference; LOESS, local regression.

Appendix - Ratio of instrument-defined MID estimate to scoring algorithm range for EQ-5D-3L country-specific tariffs and previously published instrument-defined MID estimates based on EQ-5D-5L tariffs.

Country	EQ-5D	Mean MID	Mean MID	Range of	MID	
	Version	estimate (SD)	estimate (SD)	value	estimate (all	
		based on all	after exclusion	set/tariff	transitions)	
		transitions	of maximum-		divided	
			valued scoring		by tariff	
			parameters		range	
Denmark	EQ-5D-3L	0.152 (0.032)	0.060 (0.009)	1.624	0.094	
Germany	EQ-5D-3L	0.117 (0.036)	0.060 (0.021)	1.207	0.097	
Japan	EQ-5D-3L	0.097 (0.021)	0.064 (0.012)	1.111	0.087	
Netherlands	EQ-5D-3L	0.124 (0.026)	0.073 (0.015)	1.329	0.093	
Spain	EQ-5D-3L	0.160 (0.031)	0.093 (0.011)	1.654	0.097	
Zimbabwe	EQ-5D-3L	0.105 (0.015)	0.062 (0.010)	1.145	0.092	
UK	EQ-5D-3L	0.149 (0.028)	0.082 (0.032)a	1.594	0.093	
USA	EQ-5D-3L	0.106 (0.019)	0.040 (0.026)a	1.109	0.096	
Canada	EQ-5D-5L	0.056 (0.011)b	0.037 (0.001)b	1.097	0.051	
China	EQ-5D-5L	0.069 (0.007)b	0.058 (0.005)b	1.391	0.050	
Spain	EQ-5D-5L	0.061 (0.008)b	0.045 (0.009)b	1.216	0.050	
Japan	EQ-5D-5L	0.048 (0.004)b	0.044 (0.004)b	1.025	0.047	
England	EQ-5D-5L	0.063 (0.013)b	0.037 (0.008)b	1.281	0.049	
Uruguay	EQ-5D-5L	0.063 (0.019)b	0.040 (0.010)b	1.264	0.050	

Notes: Range of value set based on published range where available or the difference in EQ-5D-3L and EQ-5D-5L index score for the worst possible health state (33333 or 55555) and the best possible health state (11111). EQ-5D-3L indicates EuroQol five-dimension threelevel instrument; EQ-5D-5L, EuroQol five-dimension five-level instrument; MID, minimally important difference; SD, standard deviation. ^aEstimates previously published by Luo *et al.* [9] ^bEstimates previously published by McClure *et al.* [15]