Estimating the minimally important difference for the EQ-5D-5L and EORTC QLQ-C30 in cancer

Bryan Bennett[1], Tara Li[2], Alexandra Soare[2], Sian Bissell O'Sullivan[2], Yemi Oluboyede[2], James W. Shaw[1], Julia Braverman[1], Siobhan Bourke[2], Louise Longworth[2]

Institutions

[1]Bristol Myers Squibb [2] PHMR Ltd

Abstract

Introduction: The minimally important difference (MID), also referred to in the literature as the (minimal) clinically important difference, can be defined as the smallest change in the score of a patient-reported outcome measure (PROM) or clinical outcome that patients perceive as important. MIDs can provide a useful tool to interpret the magnitude of changes measured using PROMs. While the EQ-5D-5L has now become a well-established PROM, limited MID estimates for the EQ-5D-5L utility index are currently available in cancer. This study aims to fill this gap in the literature by estimating within and between group-level MIDs for the EQ-5D-5L in people with cancer. Additionally, MIDs are also estimated for the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30).

Methods: Anchor-based and distribution-based methods are the most frequently applied methods to estimate MIDs. We utilised both the anchor-based methods complemented by distribution-based method as recommended to assess the variability of the estimates. Both patient-reported anchors (primary), namely: the EQ-visual analogue scale (EQ-VAS), item 29 (overall health) and item 30 (overall quality of life) of the EORTC QLQ-C30, and a clinical anchor (secondary), the Eastern Co-operative Oncology Group (ECOG) performance status, were used. Anchors for the EQ-5D-5L utility index and for each EORTC QLQ-C30 scale were selected based on the correlation between each anchor and PROM pair. Clinical change groups were defined a priori based on the minimally important change threshold for each selected anchor. Anchor-based MIDs were estimated via mean change method and linear regression for improved and deteriorated groups using patient-reported anchors. Distribution-based MIDs were determined by multiple statistics that are commonly used in the literature: 0.3 and 0.5 of the standard deviation, and the standard error measurement method.

Data source: The Cancer 2015 dataset is a longitudinal and perspective cohort. The dataset contains clinical, molecular pathology, demographic, socio-economic, and health-related quality of life (HRQoL) data of 1,685 patients, with an average follow-up of 12 months, with various newly diagnosed cancers in Victoria, Australia. Clinical data included, cancer stage, and the ECOG performance. The PROM data included the EQ-5D-5L and EORTC QLQ-C30. The cohort included patients with the most common solid-tumour cancers such as breast, lung, colorectal and prostate cancer, but also patients with more rare cancers such as head and neck, bone/soft tissue, renal or bladder cancer. Data was collected at baseline, 3-, 6- and 12-months post-consent, and continued every 12 months thereafter.

Conclusions: In the EQ-5D-5L, the resulting mean anchor-defined MID estimates were 0.02 to 0.03 for improvement and -0.04 to -0.03 for deterioration. For the EORTC QLQ-C30, changes of at least 3.55 units on the physical functioning scale, 6.42 on the role functioning scale, 4.44 on the emotional function, and 5.41 on the social functioning scale were required to constitute meaningful improvement change. The highest MID improvement needed on the symptom scale was 5.50 on the insomnia scale with the lowest MID improvement being 1.86 units on the nausea and vomiting scale, respectively. A negative change was estimated to be at least 3.77 units on the physical scale, 6.83 on the role functioning, 3.41 on the emotional, and 5.58 on the social functioning scale were required to constitute meaningful improvement change. The largest MID decrement needed on the symptom scale was 5.70 on the insomnia scale with the lowest MID decrement being 2.17 units on the nausea and vomiting scale, respectively. By utilising a cancer-specific panel dataset, (typically clinical trial or cross-sectional datasets have been used in the literature) in the estimation of MID for EQ-5D-5L and EORTC QLQ-C30, the MID estimates will be more applicable to the real-world population. Secondly, by including the patient-reported anchors in the determination of MIDs, this methodology meets what is now considered the gold standard approach.

Introduction

Cancer and the treatment of cancer has a high symptom burden on patients and can have a significant impact on health-related quality of life (Zeng et al. 2012). Health Related Quality of Life (HRQoL) is an important measurement for cancer patients, it can be used to assess the effectiveness of the treatment and identify if patients experience meaningful improvements or deterioration based on self-assessment. However, the magnitude of change in HRQoL for patients to experience that relates to a meaningful improvement or deterioration has not been thoroughly assessed.

The minimal important difference (MID) has been defined as the smallest difference in the score of a patient-reported outcome measure (PROM) or clinical outcome that is meaningful to a patient (Revicki et al. 2008). There is currently a lack of consensus in the literature regarding the terminology of MID (Houchen-Wolloff and Evans 2019). MID may refer to a change that individuals can detect, or it may refer to a difference in clinical outcome measures or PROM which lead to a meaningful change to the patient, or which results in a meaningful reduction in an important adverse outcome (Houchen-Wolloff and Evans 2019). Furthermore, the terms MID and minimal clinically important difference (MCID) are used interchangeably in the literature and may be defined differently by different authors (Cook 2008). Subsequently, we defined MID as the smallest difference in the score of a PROM or clinical outcome measure that is meaningful to a patient (Revicki et al. 2008).

HRQoL is a subjective analysis of the impact of physical and emotional health status on a patients quality of life, commonly assessed through validated PROMs (Yin et al. 2016). Two of the most commonly used PROMs in cancer are the EQ-5D-5L, a well-established generic preference-based instrument in health outcomes research and, the EORTC QLQ-C30, a cancer-specific PROM. However, a statistically significant change in a PROM score issued at two time points may not reflect a meaningful change in HRQoL that is valued by patients. Furthermore, the sample size of clinical trials is calculated to ensure that they are powered to detect improvements in primary clinical end-points but are not commonly powered to detect statistically significant changes in secondary outcomes, such as HRQoL outcomes (O'Sullivan et al. 2005). One of the advantages of the MID is that it can be applied in an informative way for clinicians, researchers, and healthcare decision agencies because it allows for an assessment and measurement of improvement (or deterioration) in HRQoL that is meaningful to the patient. Specifically, MIDs can provide a useful

tool to interpret the magnitude of changes measured using HRQoL instruments by measuring health improvement which is independently attached to a clinically meaningful measure (Mouelhi et al. 2020). From the perspective of healthcare decision agencies, the MID allows better understanding of the magnitude of the treatment effect (Johnston et al. 2015) and can provide evidence that a treatment effect is meaningful beyond statistical significance. The MID can help meet the high level of evidence required to show that improvement to HRQoL has been achieved by the technology under appraisal (Houchen-Wolloff and Evans 2019).

Currently, there is no consensus on the most appropriate method to estimate MIDs as MID estimates for a condition may vary greatly based on many factors (Kang et al. 2021) such as patient population, disease severity, follow-up time periods, and different natural histories of the disease experienced by patients. Three main approaches commonly used in the literature to estimate the MID are the anchor-based approach, the distributional approach, and the instrument-defined approach (Coretti et al. 2014). The anchor-based approach uses clinically relevant external indicators (anchors), to link changes in HRQoL in domains which are correlated to the change in the anchor measure. The change in score in the instrument of interest associated with a change in the anchor measure is the MID (Revicki et al. 2008). The distribution-based approaches including effect size and the standard error of measurement (SEM) use the distribution (standard deviation) of HRQoL scores to estimate an MID (Pickard et al. 2007). The instrument-defined approach uses the instrument for which the MID is being estimated as an internal anchor, rather than an external anchor, to estimate an MID (Xu and Cheung 2020).

Drug regulatory authorities in the US and Europe have issued guidance on the evidence they request to support the definitions of MID in the analysis of clinical trials data presented to them to support marketing authorisation applications for drugs. The EMA guidelines state that statistically significant PROM score per se are not sufficient for specific claims without demonstration of a clinically relevant treatment benefit (Committee for Medicinal Products for Human Use (CHMP) 2016). Similarly, the FDA guidelines state that distribution-based approaches should not be the main method to calculate MID (Food and Administration 2009). Concern surrounding the use of the distribution-based approach is also found in the literature where the method is criticised because the MID is not related to a clinically meaningful outcome or a patient-reported change in outcome, and therefore does not provide a measure of value to the change detected, because it does not incorporate the patient perspective (Carrasco-Labra et al. 2021) Generally, it is

recommended to generate MID estimates using different methodologies and statistical methods, and to triangulate a change in score that relates to a deterioration or improvement of the individual's condition. While there is significant literature on MID estimates for the EORTC QLQ-C30, there has been limited research into MID estimates for the EQ-5D-5L utility index in cancer. It is essential to estimate the MID for both measures in cancer patients as we can gain insights in identifying which treatments have the greatest impact on the domains of HRQoL that are most meaningful to the patient by identifying changes that are important to the patient. The use of real-world data rather than data from a clinical trial, allows us to estimate the real-word MID rather than the MID for the intervention under investigation (Coretti et al. 2014). In addition, the longitudinal nature of the data allows us to assess the responsiveness of HRQoL measures over time (Revicki et al. 2008). This study aims to determine the MID of both the EQ-5D-5L and the EORTC QLQ-C30 in a cohort of Australian cancer patients using multiple MID methodologies.

Methods

Data source and study population

The Cancer 2015 dataset is a longitudinal and prospective cohort which contains clinical, molecular pathology (i.e., tumour sample DNA), health resources, and HRQoL data of 1,685 patients with various newly diagnosed cancers in Victoria, Australia. The cohort includes patients with the most common solid-tumour cancers such as breast, lung, colorectal and prostate cancer, as well as patients with more rare cancers such as head and neck, bone/soft tissue, renal or bladder cancer. Patients' follow-up was performed at, 6- and 12-months post-consent, and continued every 12 months thereafter.

The Cancer 2015 dataset collected socio-demographic details of patients such as age, gender, country of birth, education level, and marital status. In addition, clinical data includes cancer stage, treatment summary, the Eastern Co-operative Oncology Group (ECOG) performance status, which evaluates patients' capabilities (Oken et al. 1982), and the Charlson Comorbidity Index , which predicts the risk of death within 1 year of hospitalisation for patients with specific comorbid conditions(Charlson et al. 1987). To measure the HRQoL of patients, the EQ-5D 3L and 5L, and the EORTC QLQ-C30 were administered. However, only the EQ-5D-5L and EORTC QLQ-C30 data were used for the analysis of this study. Data was collected at baseline (diagnosis) and at various follow-up points (6, 12, 24 and 36 months).

HRQoL instruments

EQ-5D-5L

EQ-5D-5L is a widely used generic, preference-based measure that assesses patients' HRQoL on 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), each being characterised by five levels of severity (no problems, some problems, moderate problems, severe problems, and extreme problems/unable to). Each health profile described by the EQ-5D-5L has a utility value usually anchored on a scale at 0 and 1, where 1 represents full health and 0 indicates a state as bad as being dead. Utility values can also be negative which suggest a state worse than being dead. The newly developed Australian value set was used to value the EQ-5D-5L utility scores. The second component of the EQ-5D-5L is the EQ-VAS which asks respondents how their general health is at the time they complete the questionnaire. Based on their own judgement, respondents self-report their health on a scale ranging from 0, which indicates 'the worst health you can imagine' and 100, which represents 'the best health you can imagine'. The EQ-VAS was used as one of the patient-reported anchors when applying the anchor-based method to estimate the MIDs, described below.

EORTC QLQ-C30

EORTC QLQ-C30 is a cancer-specific questionnaire which contains 30 items that evaluate five functional domains (physical, emotional, social, role, and cognitive), 8 symptoms (fatigue, pain, nausea/vomiting, constipation, diarrhoea, insomnia, dyspnoea, and appetite loss), the global health status and quality of life, and the financial impact of cancer. All items have a four-point scale (not at all, a little, quite a bit, and very much), except for the two global health status and QoL items. Scores for each of the 15 dimensions are converted to a 0 and 100 scale. Higher scores indicate poorer symptoms on the symptom scales, whereas for all other items, higher scores suggest better functioning and QoL. For consistency in interpretation in the analysis, the symptom scale was coded in reverse to follow the functioning scales' interpretation, i.e., 0 represents the worst possible score and 100, the best possible score. Symptom scales reported in the demographic tables are reported as standard for the EORTC QLQ-C30 where higher scores indicate higher symptom burden.

Statistical based method

Descriptive statistics were used to summarise patients' information. The primary method for the estimation of the MID in this study was the anchor-based approach to determine the meaningful

change of EQ-5D-5L utility and EORTC QLQ-C30 scores. This methodology links a change in outcome to the patient-perspective (Houchen-Wolloff and Evans 2019) which provides a value to the change observed (Carrasco-Labra et al. 2021). The distributional method was also used to provide complimentary evidence to validate the MID estimated using the anchor-based approach.

The anchor-based approach

Anchor-based methods link HRQoL to known clinically relevant indicators or to the patient determined rating of change. The anchor-based approach harnessed an external indicator (or anchor) to assign patients into different clinical change groups (CCG). These groups for each anchor were determined a priori, and they reflected various levels of change, defined as small positive change, small negative change, and no change. If patients experienced change scores on the anchors that were either lower or higher than the above pre-established thresholds they were excluded from the analysis, as these patients were not considered to have experienced a 'minimally important' change (i.e., they either experienced a change that was too low to be considered important or too high to be considered 'minimal').

The literature strongly recommends using multiple independent anchors and examining responsiveness across multiple samples. Currently, it is unclear whether an objective clinical outcome measure is more appropriate as an external anchor compared to a patient-reported anchor. The FDA has highlighted concerns that the use of a global health question, usually used as an external anchor from a HRQoL instrument, may introduce recall and response shift bias (Food and Administration 2009). However, the use of patient reported anchors such as the Patient Global Impressions scale – Change, Improvement, Severity (PGI-C, PGI-I, PGI-S) is often used and was highlighted as a major strength of the study(Hui et al. 2015). Furthermore, clinical outcome measures may not identify changes that are important to the patient (Sedaghat 2019). Many studies recommend using several anchors, and then triangulating on a single value or a small range of values to increase the robustness of the MID estimates (Yost and Eton 2005; Revicki et al. 2008) (Coon 2016). Four external indicators were chosen to be included in the analysis: one clinical (ECOG) and 3 patient-reported (EQ-5D-5L VAS, Q29 and Q30 of the EORTC QLQ-C30). Each of these anchors met the criteria for selection; all are relevant, have a clear medical interpretation, and are accepted by clinicians. Additionally, the EQ-VAS and items 29 and 30 of the EORTC QLQ-C30 have been validated in patients.

ECOG

The Eastern Cooperative Oncology Group measure (ECOG) is a performance status measure that is routinely used by clinicians to evaluate the progression of disease and the extent to which it affects patients' daily life. ECOG has five grades of severity that range from 0 and 5, where 0 indicates a patient is 'fully active', 4 suggests a patient is 'completely disabled', and 5 represents death. The CCG for the ECOG measure was defined as a change score of 1 grade on the grade scale from 0 and 5, representing the 'minimal' expected change.

EQ-VAS

The CCG threshold for the EQ-VAS, the 'minimal' expected change, was defined as a change in score of 7-10 points on the scale from 0 and 100, based on findings from previous work (Pickard et al. 2007).

Items 29 and 30 of the EORTC QLQ-C30 questionnaire

The final selected anchors used in this study were the questions relating to overall health and overall QoL from the EORTC QLQ-C30. Both these anchors are taken directly from the EORTC QLQ-C30 (questions 29 and 30) where patients can rate their overall health and QoL on a scale of 1 (very poor) to 7 (excellent). For items 29 and 30 of the EORTC QLQ-C30 questionnaire, the 'minimal' expected change was defined by a change in score of 1 point on the scale from 1 and 7.

Statistical analysis

The empirical relationship between each anchor and both the EQ-5D-5L and EORTC QLQ-C30 instruments using correlation analysis was assessed. For the base case analysis for the EQ-5D-5L, each anchor was correlated to the EQ-5D-5L utility index score, whereas for the EORTC QLQ-C30, each anchor was correlated to each EORTC QLQ-C30 scale, excluding the financial scale as it was unlikely that a correlation would exist (Musoro et al. 2020). Following Musoro et al. (2020), we calculated the correlation of both the cross-sectional utility scores to the anchor and the change correlation scores and change in anchor score for additional complementary information. To be suitable and interpretable for the estimation of an MID, each anchor must correlate at least moderately ($r > |0.30|$) with the HRQoL(Mouelhi et al. 2020). The MID estimates were triangulated on a single value or a small range of values where more than 1 anchor was found to be suitable using the correlation analysis results. To triangulate multiple MID estimates derived from different anchors a correlation-weighted average was used (Trigg and Griffiths 2021)). This allows for an

increased weighting given to stronger anchor measure found in the study. Separate triangulation of the results will be estimated for the between-group differences and within-group changes.

This study utilised the change difference and regression analysis methods. Using the change difference method, the MID was calculated as the difference in HRQoL index score change between two timepoints between the "small change" and "no change" groups that were pre-defined by the anchor. Specifically, for a given HRQoL instrument and its corresponding anchor, the MID was calculated as the difference in the average HRQoL index score between the "small positive change" and "no change" CCGs for improvement, and between the "small negative change" and "no change" CCGs for deterioration.

The regression analysis method quantified MID as the coefficient of the CCG group indicator obtained from fitting 2 regression models, one estimating improvement and one estimating deterioration with the change in HRQoL index score as the dependent variable, and the anchor as an explanatory variable (Nolan et al. 2016; Musoro et al. 2020). The CCG group indicator was coded as 0 for "no change", and 1 for "small positive change" for the improvement model, and 0 for "no change", and 1 for "small negative change" for the deterioration model. Potential confounding factors, such as patient's demographic and clinical characteristics (e.g., patient's age, sex, primary cancer site and cancer stage), were added to the regression models. The regression models were fitted for the whole sample.

The distribution-based approach

The distribution-based approach considers only the statistical distribution of HRQoL scores. It relies on the dispersion of patient's HRQoL scores to quantify how much of a change in HRQoL scores may be important for patients. The standard deviation (SD) and standard error of measurement (SEM are the two parameters mostly used in the literature(Mouelhi et al. 2020; Ousmen et al. 2018). Based on the SD, the MID was calculated as a multiplier times the SD of the HRQoL score at baseline in the patient group. Commonly 0.3 and 0.5 are the multipliers that are used in the literature in studies that have estimated MIDs using the distribution-based approach(Mouelhi et al. 2020). SEM is calculated as $SD * \sqrt{1-r}$, where r is the reliability coefficient for the HRQoL instrument. The r for the EORTC QLQ-C30 is estimated as 0.85 (Hjermstad et al. 1995), and for EQ-5D-5L index, it is estimated as 0.85 (Long et al. 2021). A threshold of 1 SEM has been widely used to determine MID (Musoro et al. 2020; Quinten et al. 2019).

Results

This paper calculated and compared the MID for the EQ-5D-5L and EORTC QLQ-C30 from a prospective population based longitudinal cancer cohort, that had collected data using both instruments to assess HRQoL.

EQ-5D-5L analysis

Table 1 presents a summary of demographic and clinical characteristics of patients who were in the EQ-5D-5L sample. The mean age of participants at baseline (N= 464) was 63 years old, and 56% were female. At baseline there were 103 unique EQ-5D-5L health states reported with an average EQ-5D-5L index score of 0.90 ± 0.14. 26% of the sample reported a perfect health state utility at baseline.

Table 1. EQ-5D-5L sample demographic and clinical characteristics

Patient characteristics N=464		
	Mean (SD)	**95% Confidence Interval**
Age	63 (11)	62 - 64
EQ-5D-5L score (baseline)	0.9042 (0.14)	0.8918 - 0.9166
EQ5D VAS (baseline)	80.01 (15.86)	
	n	**% of total N of patients**
Number of health states (baseline)	103	
Health state 11111	*122*	*26 %*
Number of follow-ups		
1	27	6%
2	70	15%
3	117	25%
4	177	38%
5	73	16%

Sex		
Male	204	44%
Female	260	56%
Cancer site at first diagnosis		
Breast	161	35%
Genitourinary	106	23%
Head and Neck	63	14%
Colorectal	60	13%
Lung	25	5%
All others	46	10%
Missing	3	0.7%
Staging		
Unknown	8	2%
Distant metastases	12	3%
Regional lymph nodes	108	23%
Invasion of adjacent tissue or organs	44	9%
Localised to the tissue of origin	262	56%
Missing	30	6%

A total of 4 potential anchors were initially assessed for both the EQ-5D-5L and EORTC QLQ-C30 scales, of which one was a clinical measure (ECOG) and 3 were patient-reported anchors (EQ-VAS, Q29 and Q30 of the EORTC QLQ-C30). Table 2 provides estimates of cross-sectional correlations between the EQ-5D-5L scores and each anchor, and the correlation between their change scores. The cross-sectional correlations between HRQoL scales and anchors ranged from -0.10 to 0.6,

while the correlations between their change scores ranged from -0.06 to 0.4. Examining each anchor, we observed change score correlations of -0.06 for ECOG, 0.41 for VAS, 0.34 and 0.32 for Q29 and Q30 in the EORTC QLQ--C30.

Table 2. Correlations between anchors and EQ-5D-5L score

	Cross sectional	Change score
	Correlation	**Correlation**
ECOG	-0.10	-0.06
VAS	**0.60**	**0.41**
Q29	**0.57**	**0.34**
Q30	**0.57**	**0.32**
Numbers highlighted in **bold** meet the threshold (r >| 0.30|) for inclusion for MID estimation		

Figure 1 examines the strength of the association between the change in anchor score and the change in EQ-5D-5L index score. The dotted lines represent the change in anchor for a deterioration or improvement. The correlations for both the cross-sectional and change scores between the clinical anchor, ECOG, and the EQ-5D-5L score were less than the 0.3 threshold so ECOG was not retained for further analysis.

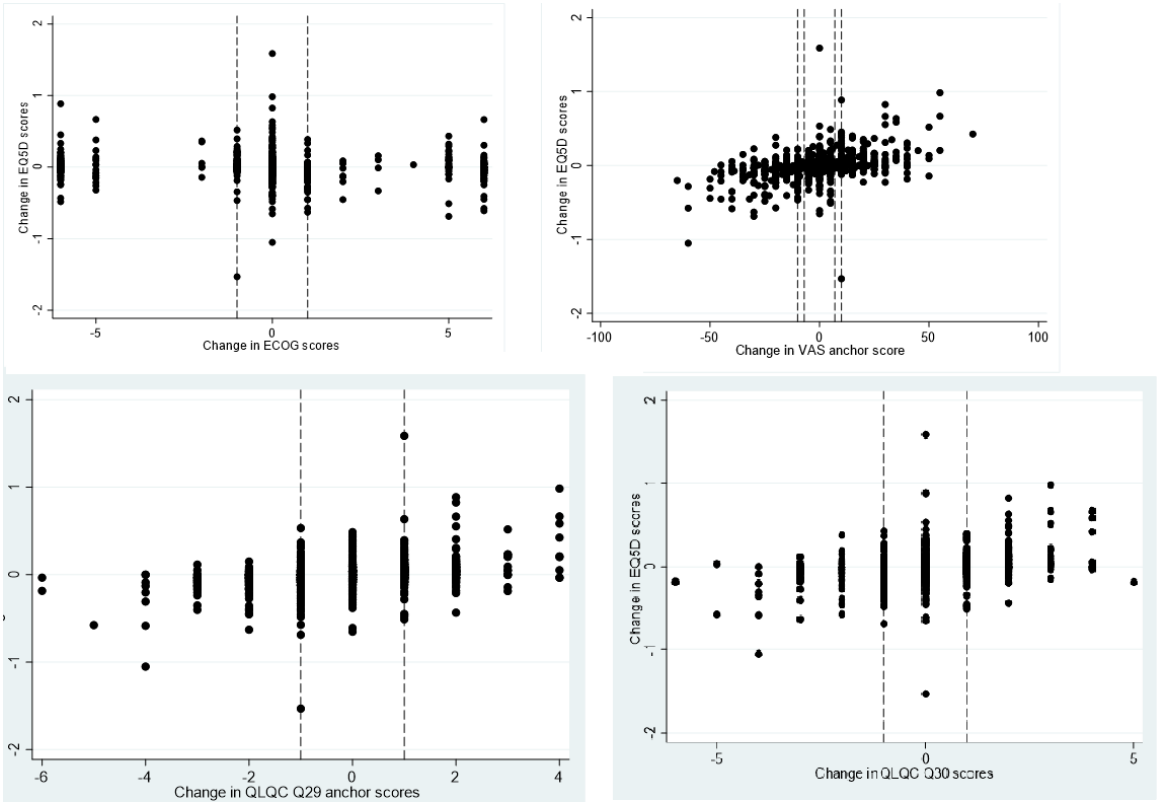

Figure 1. Change in EQ-5D-5L score and anchor

Results in Table 3 present the MID estimates for those anchors that satisfied the anchor threshold criteria. The MID estimates varied according to the scale, direction of change scores (improvement

versus deterioration), and anchor. Within our analysis the MID estimates were in the expected direction, where positive was associated with an improvement and negative scores was associated with a deterioration in utility score, respectively. The MID estimate of the mean change anchor-based method (within group change) ranged from 0.01 to 0.03 in the improvement category and -0.03 to -0.04 in the deterioration category. In the regression-based method (between group change) this ranged from 0.02 to 0.06 in the improvement category and -0.04 to -0.03 in the deterioration category. Table 3 also describes the MID estimates for the distribution approach where the results of the MID analysis ranged from 0.04 to 0.07. Only the mean change improvement results of the Q30anchor based analysis did not satisfy Cohen's recommendation that it should be considered a MID.

Table 3. EQ-5D- 5L MID mean change and linear regression

	Anchor-based approach				Distribution-based approach		
	Mean change		Regression model				
	Improve	**Deteriorate**	**Improve**	**Deteriorate**	**0.5**	**0.3**	**SEM**
ECOG	n/a	n/a	n/a	n/a	n/a	n/a	n/a
VAS	0.02 **(0.2)**	-0.04 **(-0.3)**	0.02 **(0.2)**	-0.04 **(-0.3)**	0.07	0.04	0.06
Q29	0.03 **(0.2)**	-0.03 **(-0.2)**	0.06 **(0.4)**	-0.03 **(-0.2)**	0.07	0.04	0.06
Q30	0.01 **(0.1)**	-0.04 **(-0.3)**	0.05 **(0.3)**	-0.03 **(-0.2)**	0.07	0.04	0.06
Weighted EQ-5D MID	0.02 **(0.2)**	-0.04 **(-0.3)**	0.04 **(0.3)**	-0.03 **(-0.2)**	n/a	n/a	n/a
Bold figures reflect effect size (ES)							

Table 3 also presents a single triangulated summary of MIDs based on a correlation-weighted average following the method given in (Trigg and Griffiths 2021). For improvement the mean change method (within group change) resulted in an improvement of 0.02 and a deterioration of -0.04. For the regression model we triangulated a summary improvement of 0.04 and a deterioration of -0.03.

EORTC QLQ-C30

Table 4 presents a summary of demographic and clinical characteristics of patients who were in the EORTC QLQ-C30 sample. The mean age of participants at baseline (N= 799) was the same for both samples at 63 years, and 55% were female. The most common type of cancer reported was

breast cancer (35%) followed by genitourinary cancer (22%). The EORTC QLQ-C30 sample had more respondents who had distant metastases (6%) in comparison to the EQ-5D-5L sample (3%). Baseline symptoms were assessed and reported using the EORTC QLQ-C30 outcome measure. As shown in Table 4, self-rated function in our sample was relatively high. Across subscales, sample mean scores ranged between 81.5 and 85.2 (on a 100-point scale).

Table 4. EORTC QLQ-C30 sample demographic and clinical characteristics

Patient characteristics		
	Mean	**Standard deviation (SD)**
Age	63	12
EORTC QLQ-C30 baseline		
Overall health	5.35 (1.22)	1.22
Quality of life	5.54	1.25
Global health status (item 29)/QoL scale (item 30)	74.08 (19.67)	19.67
Physical functioning	85.2	18.2
Role functioning	81.9	25.6
Emotional functioning	81.5	20.1
Cognitive functioning	83.3	19.8
Social functioning	82.5	25.1
Pain	18.7	24.3
Dyspnea	13.9	22.8
Appetite loss	10.6	22.1
Constipation	9.9	19.8
Diarrhoea	7.6	17.5
Fatigue	26.3	23.0
Nausea/Vomiting	4.4	11.4
Financial problems	13.7)	25.41
Number of follow-ups	**n**	**% of total N of patients**
1	45	6%
2	154	19%
3	219	27%
4	206	26%
5	109	14%
6	32	4%
7	15	2%
8	9	1%
9	3	0.4%
10	2	0.3%
Sex		
Male	357	45%

Female	442	55%
Cancer type		
Breast	278	35%
Genitourinary	178	22%
Head and Neck	95	12%
Colorectal	101	13%
Central nervous system	8	1%
Lung	45	6%
All others	91	11%
Missing	3	0.38%
Staging		
Unknown	16	2%
Distant metastases	50	6%
Regional lymph nodes	188	24%
Invasion of adjacent tissue or organs	70	9%
Localised to the tissue of origin	437	55%
Missing	38	5%

The same four potential anchors, namely the clinical (ECOG) and 3 patient-reported anchors (VAS, Q29 and Q39 of the EORTC QLQ-C30) were tested for suitability for inclusion for MID estimation. Table 5 provides the correlations between the EORTC QLQ-C30 and anchors and the correlation between the changes in EORTC QLQ-C30 and in the anchors. The cross-sectional correlations between HRQoL scales and anchors ranged from 0.11 to 0.67 in absolute value, while the correlations between their change scores ranged from 0.20 to 0.44. For the absolute value it was determined that the functional scales physical, role, and social functioning were correlated at least moderately ($|r| > 0.30$) with each of the health anchors selected. Generally, patient-reported anchors showed higher correlations with HRQoL scales compared to clinical anchors. In addition, the patient reported anchors met the correlation threshold for estimation of MID for more item scales compared to clinical measures. For the relative change score value, no scale value of the EORTC QLQ-C30 for the ECOG anchor met the correlation threshold for inclusion. The correlations highlighted in bold were retained for estimation of MID values.

Table 5.Correlations between EORTC QLQ-C30 and anchors

		Cross sectional	Change score
Scale	**Anchor**	**Correlation**	**Correlation**
PF	ECOG	**-0.48**	-0.27
RF	ECOG	**-0.41**	-0.20
EF	ECOG	-0.17	-0.09
CF	ECOG	-0.24	-0.12
SF	ECOG	**-0.35**	-0.14
FA	ECOG	**-0.38**	-0.22
NV	ECOG	-0.24	-0.14
PA	ECOG	-0.25	-0.15
DY	ECOG	-0.25	-0.10
SL	ECOG	-0.17	-0.07
AP	ECOG	**-0.30**	-0.14
CO	ECOG	-0.17	-0.08
DI	ECOG	-0.12	-0.04
QL	ECOG	**-0.34**	-0.20
PF	VAS	**0.61**	**0.40**
RF	VAS	**0.64**	**0.39**
EF	VAS	**0.49**	0.29
CF	VAS	**0.47**	0.26
SF	VAS	**0.58**	**0.36**
FA	VAS	**0.66**	**0.44**
NV	VAS	**0.36**	0.22
PA	VAS	**0.56**	**0.33**
DY	VAS	**0.43**	0.24
SL	VAS	**0.40**	0.20
AP	VAS	**0.43**	0.28
CO	VAS	0.27	0.11
DI	VAS	0.20	0.07
QL	VAS	-0.11	**0.57**
PF	Q29	**0.60**	**0.40**
RF	Q29	**0.63**	**0.40**
EF	Q29	**0.49**	0.28
CF	Q29	**0.46**	0.24
SF	Q29	**0.57**	**0.34**
FA	Q29	**0.67**	**0.43**
NV	Q29	**0.37**	0.23
PA	Q29	**0.58**	**0.35**
DY	Q29	**0.44**	0.23
SL	Q29	**0.42**	0.20
AP	Q29	**0.44**	0.29

CO	Q29	0.27	0.10
DI	Q29	0.23	0.11
QL	Q29	n/a	n/a
PF	Q30	**0.58**	**0.38**
RF	Q30	**0.63**	**0.39**
EF	Q30	**0.52**	**0.31**
CF	Q30	**0.46**	0.22
SF	Q30	**0.62**	**0.39**
FA	Q30	**0.64**	**0.38**
NV	Q30	**0.37**	0.23
PA	Q30	**0.55**	**0.33**
DY	Q30	**0.44**	0.19
SL	Q30	**0.42**	0.18
AP	Q30	**0.43**	0.29
CO	Q30	0.25	0.10
DI	Q30	0.22	0.10
QL	Q30	n/a	n/a

*PF=physical function, RF= role functioning, EF= emotional functioning, CF= cognitive functioning SF= social functioning, FA= fatigue, NV= nausea vomiting, PA= pain, DY= dyspnoea, SL= insomnia, AP= appetite loss, CO = constipation, DI = diarrhoea, QL = Global health status / QoL

Bold figures satisfy the threshold (|r| > 0.30) for MID estimation using change correlation method

Results in Table 6 present the MID estimates for the anchors that satisfied the anchor threshold criteria. As before, the MID estimates were in the expected direction with positive scores within the improvement category and negative scores in the deterioration category. The ECOG anchor was only suitable for estimating the MID for physical functioning, relationship functioning, social functioning, fatigue, and global quality of life. The MID estimate using the mean change anchor-based method (within group change) for physical functioning ranged from 2.85 to 5.94 in the improvement category and -1.17 to -6.20 in the deterioration category. For role functioning the improvement ranged from 3.0 to 8.66 and deterioration from -5.45 to -7.61. Further MID estimates for the additional scales are described in table 6. In the regression method (between group change) the MID improvement ranged from 1.83 to 9.25 and deterioration from -2.12 to -8.12. Table 6 also describes the MID estimates for the distribution approach where the results of the MID analysis ranged from 3.52 to 13.63.

Results of the triangulation is presented in Table 6. A range of summary MIDs based on a correlation-weighted average was generated for 11 of the 14 item scales of the QLQ-C30. The MIDs for most EORTC QLQ-C30 scales ranged from 2.02 to 6.83 points in absolute values for the mean change method and for the regression results ranged between 1.68 to 7.11. Two scales

failed to satisfy the minimum ES needed to be considered a true MID, the improvement scores for both cognitive functioning and dyspnoea.

Table 6.EORTC QLQ-C30 MID estimates

	Mean change		Regression model		Distribution approach		
Scale	**Improve**	**Deteriorate**	**Improve**	**Deteriorate**	**0.5**	**0.3**	**SEM**
PF	2.85 to 5.94 **(3.55)**	-6.20 to -1.27 **(-3.77)**	2.75 to 6.37 **(3.64)**	-6.58 to -3.01 **(-4.28)**	5.37-5.4	7.59-7.64	7.59-7.64
RF	5.63 to 8.66 **(6.42)**	-7.60 to -5.57 **(-6.83)**	5.39 to 9.25 **(6.31)**	-8.12 to -6.43 **(-7.11)**	13.63-13.67	8.2	11.57-11.8
EF	4.04 to 4.87 **(4.44)**	-4.82 to-2.18 **(-3.56)**	4.39 to 5.08 **(4.65)**	-4.96 to -1.96 **(-3.41)**	10.72	6.43	9.09
CF	1.97 to 2.67 **(2.27)** †	-3.29 to-2.99 **(-3.15)**	2.18 to 2.68 **(2.44)** †	-3.13 to -2.90 **(-3.03)**	10.11	6.06	8.58
SF	4.91 to 5.83 **(5.42)**	-7.19 to -4.21 **(-5.68)**	4.98 to 5.81 **(5.41)**	-7.04 to -4.12 **(-5.56)**	12.93-12.95	7.77	10.97-10.99
FA	4.62 to 7.37 **(5.64)**	-7.61 to-6.50 **(-6.81)**	4.95 to 7.23 **(5.76)**	-8.27 to -6.47 **(-6.93)**	11.48-11.5	6.89	9.76
NV	1.12 to 2.89 **(1.9)**	-2.42 to -1.87 **(-2.17)**	1.16 to 2.71 **(1.86)**	-2.26 to -2.12 **(-2.19)**	5.86	3.52	4.98
PA	3.83 to 5.41 **(4.71)**	-7.62 to-5.45 **(-6.77)**	3.39 to 5.37 **(4.31)**	-7.36 to -5.48 **(-6.54)**	12	7.2	10.18
DY	0.01 to 2.68 **(1.45)** †	-4.41 to -3.55 **(-3.95)**	0.25 to 2.11 **(1.39)** †	-4.48 to -3.29 **(-4.08)**	10.49	6.29	8.9
SL	4.24 to 7.09 **(5.5)**	-5.71 to -4.89 **(-5.25)**	4.43 to 7.01 **(5.63)**	-6.18 to -5.05 **(-5.70)**	14.64	8.78	12.42
AP	3.04 to 5.21 **(3.84)**	-4.73 to -2.55 **(-3.43)**	3.00 to 4.85 **(3.84)**	-4.87 to -2.32 **(-3.42)**	11.71	7.03	9.94

*PF=physical function, RF= role functioning, EF= emotional functioning, CF= cognitive functioning SF= social functioning, FA= fatigue, NV= nausea vomiting, PA= pain, DY= dyspnoea, SL= insomnia, AP= appetite loss, CO = constipation, DI = diarrhoea

Bold figure EORTC QLQ-C30 sub-scale Triangulated MID estimates

† Failed to meet Cohen's criteria a true MID estimate ES is required to be between ≥0.2 and ≤0.8

Discussion

Estimation of MID in HRQoL assessments is important when determining the effectiveness of the treatment and identifying if patients experience meaningful improvements or deterioration based on self-assessment. In the EQ-5D-5L, the resulting mean anchor-defined MID estimates were 0.02 to 0.03 for improvement and -0.04 to -0.03 for deterioration. For the EORTC QLQ-C30, changes of at least 3.55 units on the physical functioning scale, 6.42 on the role functioning scale, 4.44 on the emotional function, and 5.41 on the social functioning scale were required to constitute meaningful improvement change. The highest MID improvement needed on the symptom scale was 5.50 on the insomnia scale with the lowest MID improvement being 1.86 units on the nausea and vomiting scale, respectively. A negative change was estimated to be at least 3.77 units on the physical scale, 6.83 on the role functioning, 3.41 on the emotional, and 5.58 on the social functioning scale were required to constitute meaningful improvement change. The largest MID decrement needed on the symptom scale was 5.70 on the insomnia scale with the lowest MID decrement being 2.17 units on the nausea and vomiting scale, respectively.

Our estimates identified improvements or deterioration in the expected direction for both the EQ-5D-5L and the EORTC QLQ-C30. For the EQ-5D-5L, the estimates for deterioration tended to be larger than those for improvement, a pattern which also carried over in large part for the EORTC QLQ-C30. This may be due to the high baseline values - more difficult to detect improvement when high and easier to detect deterioration when baseline is high. However, the current evidence on whether there is significance between the differences in the magnitude of change between deteriorating and improving scores is conflicting (Musoro et al. 2018; Maringwa et al. 2011; Cella et al. 2002).

The MIDs for HRQoL measures estimated using the distribution-based approach were somewhat consistently larger than those using the anchor-based method. In the EQ-5D-5L analysis, we found that the distribution-based and the anchor-based regression approach produced similar MID estimates, supporting the validity of the range of MID estimates. Examining the distributional estimates of the MID for the EORTC QLQ-C30, it is apparent that the estimates are higher than those obtained from the anchor-based method. Distribution-based approaches have the advantage of being easy to calculate as they do not require an external criterion, unlike the anchor-based method. However, the distribution-based approach relies on the assumption of normality particularly in healthier patient populations who produce more skewed score

distributions (Pickard et al. 2007) and they do not provide insight into the importance of the difference (de Vet et al. 2006). Additionally, a limitation of using the standard deviation to estimate MID for health state utility in the EQ-5D-5L is that the closer the average utility is to 1, the smaller the SD and thus the smaller the MID estimated.

The EQ-5D-5L MID for improvement in our study was significantly lower than previously estimated in US and UK cancer patients. In our study we estimated the MID for improvement ranging from 0.02 – 0.07, however, MID cancer estimates for the EQ-5D-3L UK index was 0.08 and for the US was 0.06 (Pickard et al. 2007). The EQ-5D- 5L utility score at baseline in the present study was0.90, with 26% of the respondents reporting perfect health. Comparing this baseline figure to previous cancer research using the same data in 2018, we see that the average tariff at baseline for the EQ-5D-3L respondents was 0.751 compared with 0.851 for EQ-5D-5L respondents (Lorgelly et al. 2018) with 23% of respondents in the 5L version reporting perfect baseline health, 3% lower than what was observed in our study. This demonstrates that there can be a large difference between baseline estimates and thus MID estimates given the tariff values used. The baseline estimated in our sample, 0.90, and the baseline in the 2018 study, 0.85, to the Australian general population, we see the mean utility was 0.91 (McCaffrey et al. 2016). This would indicate that the impact of cancer has minimal impact on the Australian population HRQoL. Therefore, we can hypothesise that because of these high levels of perfect health at baseline, improvement can occur at a lower magnitude in our sample and may explain our lower improvement estimate. An example of this can be found in the Bedard et al. where a study of cancer patients with bone metastases, patients reported a higher improvement MID for pain than previously reported elsewhere (Bedard et al. 2014). Given the sample baseline experienced higher pain in the study, the authors hypothesised that the patients had more capacity to improve given their starting pain at baseline, and thus had a higher improvement score for pain. Given that MID values can vary by disease group/severity, patient baseline status, direction of change, and demographic factors, the characteristics of our sample (relatively stable, early disease state, and high baseline HRQoL) may influence the MID generated.

It is difficult to compare to previous research because of the change of the EQ-5D version from 3L to 5L, as well as the different country specific tariffs, which can result in substantially different health utilities between countries. To our knowledge, the only MIDs for EQ-5D utility scores for cancer patients in Australia which have been previously reported relate to the 3-level version of

the EQ-5D. Here, using the 0.5 SD approach, an MID of 0.09 was estimated for cancer (Tsiplova et al. 2016), which is higher than the 0.07 MID that we estimated in our study.

Examining MIDs estimated outside of cancer for the EQ-5D-5L, a MID of 0.0917 was calculated for an elderly population with hypertension in China using an instrument-defined approach. Additionally, in stroke patients the MID estimated was a change of 0.10 in the EQ-5D index score (Tsiplova et al. 2016; Wong et al. 2020). This is considerably higher than our estimate of 0.02 – 0.07. Furthermore, in the Australian population larger MID values for Chronic obstructive pulmonary disease (COPD)(0.11), asthma (0.11), and anxiety or depression (0.11) have been estimated, which are greater than the MIDs estimated for cancer. This may suggest that the cancer patients have a greater adaption to illness and resulting response shift in HRQoL.

Previous research for participants who have advanced cancer demonstrated higher MID values compared to our study. Examining the literature for EORTC QLQ-C30 MID scores we see advanced cancer patients have a meaningful change for improvement ranging from 10.1 units (physical functioning) to 13.5 units (role functioning). Similarly in brain cancer patients (Maringwa et al. 2011), EORTC QLQ-C30 MID estimates for improvement was reported to be nine for physical functioning, 14 for role functioning, five for social functioning, and 14 for fatigue. These values are different to the estimates obtained in our study, where we observed lower values for physical functioning, fatigue, and social functioning. The higher values were attributed to patient characteristics: specifically, the higher levels of pain at baseline (Bedard et al. 2014). Our study estimated lower MID unit changes, however our sample consisted of lower numbers of advanced cancer compared to Maringwa et al. 2011. The MID estimates in our study for most scales were within the range of 5-8 unit changes this was similar to cocks et al. (Cocks et al. 2012). In patients with multiple myeloma in longitudinal data across multiple cancer sites in Norway, researchers using the response shift methodology found EORTC QLQ-C30 scores varied from 2 to 17.5 for improved patients, which is a broader range for our study (range from 2 to 8.66). The MID for deterioration in the study ranged between 12.2 to 27, which is again a larger range than reported in our study (range from -1.87 to -7.62).

Strengths

This study has followed best practice recommendations, identified from a targeted review of the literature and available guidelines from the FDA and EMA. We used multiple anchors in our study, one objective clinical measure (ECOG) and three patient reported outcomes (Q29, Q30 and the EQ-5D VAS), to ensure that the anchors used were credible and appropriate. The use of multiple

anchors helped to validate our results. In addition, there is some agreement in the literature that the use of patient-reported anchor-based approach is the optimal way to determine the MID as it directly captures the patients' preferences and is now considered the gold standard approach (Guyatt et al. 2002; King 2011).

By utilising a cancer-specific longitudinal dataset, (typically clinical trial or cross-sectional datasets have been used in the literature) in the estimation of MID for EQ-5D-5L and EORTC QLQ-C30, the MID estimates are more applicable to the real-world population and may be more relevant to reimbursement decision-making agencies. There is a lack of guidance around the most appropriate minimally important difference thresholds, however the Germany reimbursement agency Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) has highlighted that the use of MID from Pickard et al (2007) is unsuitable due to its estimation from a cross sectional design rather than longitudinal data. This highlights that the current cross-sectional MID estimates may not be suitable and fit for use in HTA applications(IQWiG 2018).

Limitations

The choice of clinical anchor for MID studies in retrospective longitudinal data is challenging. Within our study the ECOG was selected as the clinical anchor and demonstrated very little movement from the clinically stable group. The lack of change in the ECOG values resulted in correlation coefficients that were below the threshold for inclusion of ECOG as an anchor to estimate the MID in the study. We included a correlation between cross-sectional HRQoL score and ECOG status, where correlation values were above the threshold which allowed us to estimate anchor-based MID scores. However, despite this meeting the correlation threshold, there are concerns about the plausibility of the selected anchor, as well as the reliability of the estimated MIDs for this anchor.

The longitudinal nature of the Cancer 2015 cohort data posed a limitation for the study. Individuals in the dataset at earlier dates completed the EQ-5D-3L but the cohort was switched to the EQ-5D-5L for later observations and for individuals recruited later in the data set. Hence the EQ-5D-5L analysis did not use "true" EQ-5D baseline value, as there was no consistent start date (e.g., 14 days after diagnosis), compared to that of the EORTC QLQ-C30 data. In our data we took the patients first observation with EQ-5D-5L as the "observed" baseline in our dataset. Our data included 103 patients who were determined to have entered the Cancer 2015 cohort where their first observation in our data set was their "true" baseline. A simple analysis (between groups t-test) comparing the two populations (true baseline and observed baseline) was undertaken to see

if there was difference between the utility means, however no statistically significant difference (p= 0.58) was observed between the groups. Thus, we can conclude that this limitation in data did not have an impact on the analysis.

In addition, there were rarer solid cancer groups, such as head and neck cancer, where a higher prevalence was observed in our dataset. Cancer 2015 is a longitudinal prospective cohort of cancer patients treated in 5 hospitals in Victoria, Australia. The cohort successfully recruited the expected major solid such as breast, lung, colorectal and prostate cancer(Parisot et al. 2015). It also has recruited fewer common cancers such as head and neck cancer, bladder cancer and bone and soft tissue cancer. In Australia in 2009 there was 3.4% of all cancers were diagnosed as head and neck cancer(AIHW (2014)), in the Cancer 2015 cohort sample we have approximately 23% of patients with head and neck cancer, highlighting a higher prevalence in this cancer category, that could impact the generalisability of the MID estimates.

Conclusion

Knowledge of the minimal amount of change required for patients to experience a relevant improvement or deterioration is important when determining the impact of treatments on patients' HRQoL and the effectiveness of the treatment. Further studies should identify more robust clinical anchors. These studies should also focus on cancer type and severity to estimate the MID for individual cancers. Identification of robust meaningful change in HRQoL can also be used as a tool to aid researchers in the determination of the sample size required for clinical trials and determine if changes in HRQoL are important to patients.

References

AIHW ((2014)) Head and neck cancers in Australia. Canberra, ACT

Bedard G, Zeng L, Zhang L, Lauzon N, Holden L, Tsao M, Danjoux C, Barnes E, Sahgal A, Poon M (2014) Minimal important differences in the EORTC QLQ-C30 in patients with advanced cancer. Asia-Pacific Journal of Clinical Oncology 10 (2):109-117

Carrasco-Labra A, Devji T, Qasim A, Phillips MR, Wang Y, Johnston BC, Devasenapathy N, Zeraatkar D, Bhatt M, Jin X (2021) Minimal important difference estimates for patient-reported outcomes: a systematic survey. Journal of Clinical Epidemiology 133:61-71

Cella D, Bullinger M, Scott C, Barofsky I Clinical Significance Consensus Meeting G. Group vs individual approaches to understanding the clinical significance of differences or changes in quality of life. In: Mayo Clin Proc, 2002. vol 4. pp 384-392

Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. Journal of chronic diseases 40 (5):373-383

Cocks K, King M, Velikova G, de Castro Jr G, St-James MM, Fayers P, Brown J (2012) Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. European journal of cancer 48 (11):1713-1721

Committee for Medicinal Products for Human Use (CHMP) (2016) The use of patient-reported outcome (PRO) measures in oncology studies: Appendix 2 to the guideline on the evaluation of

anticancer medicinal products in man. vol EMA/CHMP/292464/2014. European Medicines Agency

Cook CE (2008) Clinimetrics corner: the minimal clinically important change score (MCID): a necessary pretense. Journal of Manual & Manipulative Therapy 16 (4):82E-83E

Coon CD Telling the interpretation story: the case for strong anchors and multiple methods. In: QUALITY OF LIFE RESEARCH, 2016. SPRINGER VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS, pp 1-2

Coretti S, Ruggeri M, McNamee P (2014) The minimum clinically important difference for EQ-5D index: a critical review. Expert review of pharmacoeconomics & outcomes research 14 (2):221-233

de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM (2006) Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health and quality of life outcomes 4 (1):1-5

Food, Administration D (2009) Patient reported outcome measures: use in medical product development to support labelling claims. Washington DC

Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Group CSCM Methods to explain the clinical significance of health status measures. In: Mayo Clinic Proceedings, 2002. vol 4. Elsevier, pp 371-383

Hjermstad MJ, Fossa SD, Bjordal K, Kaasa S (1995) Test/retest study of the European Organization for Research and Treatment of Cancer Core Quality-of-Life Questionnaire. Journal of Clinical Oncology 13 (5):1249-1254

Houchen-Wolloff L, Evans RA (2019) Unravelling the mystery of the 'minimum important difference'using practical outcome measures in chronic respiratory disease. Chronic Respiratory Disease 16:1479973118816491

Hui D, Shamieh O, Paiva CE, Perez-Cruz PE, Kwon JH, Muckaden MA, Park M, Yennu S, Kang JH, Bruera E (2015) Minimal clinically important differences in the Edmonton Symptom Assessment Scale in cancer patients: A prospective, multicenter study. Cancer 121 (17):3027-3035. doi:10.1002/cncr.29437

IQWiG (2018) Bosutinib (chronische myeloische Leukämie) – Nutzenbewertung gemäß § 35a SGB V.

Johnston BC, Ebrahim S, Carrasco-Labra A, Furukawa TA, Patrick DL, Crawford MW, Hemmelgarn BR, Schunemann HJ, Guyatt GH, Nesrallah G (2015) Minimally important difference estimates and methods: a protocol. BMJ open 5 (10):e007953

Kang M, Marts L, Kempker JA, Veeraraghavan S (2021) Minimal clinically important difference in idiopathic pulmonary fibrosis. Breathe 17 (2)

King MT (2011) A point of minimal important difference (MID): a critique of terminology and methods. Expert review of pharmacoeconomics & outcomes research 11 (2):171-184

Long D, Polinder S, Bonsel GJ, Haagsma JA (2021) Test–retest reliability of the EQ-5D-5L and the reworded QOLIBRI-OS in the general population of Italy, the Netherlands, and the United Kingdom. Quality of Life Research 30 (10):2961-2971

Lorgelly P, Cubi-Molla P, Pennington M, Norman R (2018) Comparing the EQ-5D-3L and EQ-5D-5L in a cohort of cancer patients. Value in Health 21:S226

Maringwa J, Quinten C, King M, Ringash J, Osoba D, Coens C, Martinelli F, Vercauteren J, Cleeland C, Flechtner H (2011) EORTC PROBE project and the Lung Cancer Group. Minimal important differences for interpreting health-related quality of life scores from the EORTC QLQ-C30 in lung cancer patients participating in randomized controlled trials. Support Care Cancer 19:1753-1760

McCaffrey N, Kaambwa B, Currow DC, Ratcliffe J (2016) Health-related quality of life measured using the EQ-5D–5L: South Australian population norms. Health and quality of life outcomes 14 (1):1-12

Mouelhi Y, Jouve E, Castelli C, Gentile S (2020) 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 18 (1):1-17

Musoro J, Coens C, Singer S, Tribius S, Oosting S, Groenvold M, Simon C, Machiels J, Grégoire V, Velikova G (2020) EORTC Head and Neck and Quality of Life Groups Minimally important differences for interpreting European Organisation for Research and Treatment of Cancer quality of life questionnaire Core 30 scores in patients with head and neck cancer. Head & neck 42 (11):3141-3152

Musoro JZ, Bottomley A, Coens C, Eggermont AM, King MT, Cocks K, Sprangers MA, Groenvold M, Velikova G, Flechtner H-H (2018) Interpreting European Organisation for Research and Treatment for Cancer Quality of life Questionnaire core 30 scores as minimally importantly different for patients with malignant melanoma. European Journal of Cancer 104:169-181

Nolan CM, Longworth L, Lord J, Canavan JL, Jones SE, Kon SS, Man WD (2016) The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. Thorax 71 (6):493-500

O'Sullivan AK, Thompson D, Drummond MF (2005) Collection of health-economic data alongside clinical trials: is there a future for piggyback evaluations? Value in Health 8 (1):67-79

Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, Carbone PP (1982) Toxicity and response criteria of the Eastern Cooperative Oncology Group. American journal of clinical oncology 5 (6):649-656

Ousmen A, Touraine C, Deliu N, Cottone F, Bonnetain F, Efficace F, Brédart A, Mollevi C, Anota A (2018) Distribution-and anchor-based methods to determine the minimally important difference on patient-reported outcome questionnaires in oncology: a structured review. Health and quality of life outcomes 16 (1):1-12

Parisot JP, Thorne H, Fellowes A, Doig K, Lucas M, McNeil JJ, Doble B, Dobrovic A, John T, James PA, Lipton L, Ashley D, Hayes T, McMurrick P, Richardson G, Lorgelly P, Fox SB, Thomas DM (2015) "Cancer 2015": A Prospective, Population-Based Cancer Cohort-Phase 1: Feasibility of Genomics-Guided Precision Medicine in the Clinic. J Pers Med 5 (4):354-369. doi:10.3390/jpm5040354

Pickard AS, Neary MP, Cella D (2007) Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health and quality of life outcomes 5 (1):1-8

Quinten C, Kenis C, Decoster L, Debruyne PR, De Groof I, Focan C, Cornélis F, Verschaeve V, Bachmann C, Bron D (2019) Determining clinically important differences in health-related quality of life in older patients with cancer undergoing chemotherapy or surgery. Quality of Life Research 28 (3):663-676

Revicki D, Hays RD, Cella D, Sloan J (2008) Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. Journal of clinical epidemiology 61 (2):102-109

Sedaghat AR (2019) Understanding the minimal clinically important difference (MCID) of patient-reported outcome measures. Otolaryngology–Head and Neck Surgery 161 (4):551-560

Trigg A, Griffiths P (2021) Triangulation of multiple meaningful change thresholds for patient-reported outcome scores. Quality of Life Research 30 (10):2755-2764

Tsiplova K, Pullenayegum E, Cooke T, Xie F (2016) EQ-5D-derived health utilities and minimally important differences for chronic health conditions: 2011 Commonwealth Fund Survey of Sicker Adults in Canada. Quality of Life Research 25 (12):3009-3016

Wong ELY, Xu RH, Cheung AWL (2020) Health-related quality of life in elderly people with hypertension and the estimation of minimally important difference using EQ-5D-5L in Hong Kong SAR, China. The European Journal of Health Economics 21 (6):869-879

Xu RH, Cheung AW-l (2020) 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 10 (11):e039397

Yin S, Njai R, Barker L, Siegel PZ, Liao Y (2016) Summarizing health-related quality of life (HRQOL): development and testing of a one-factor model. Population health metrics 14 (1):1-9

Yost KJ, Eton DT (2005) Combining distribution-and anchor-based approaches to determine minimally important differences: the FACIT experience. Evaluation & the health professions 28 (2):172-191

Zeng L, Chow E, Zhang L, Tseng L-M, Hou M-F, Fairchild A, Vassiliou V, Jesus-Garcia R, El-Din MAA, Kumar A (2012) An international prospective study establishing minimal clinically important differences in the EORTC QLQ-BM22 and QLQ-C30 in cancer patients with bone metastases. Supportive Care in Cancer 20 (12):3307-3313