

# Chapter 5

## Advanced Topics



This chapter provides guidance on a number of advanced topics, building on the content of earlier chapters. Our aims are

- to show how changes and differences in EQ-5D values and EQ VAS scores can be analysed;
- to discuss what a Minimally Important Difference (MID) means in the context of EQ-5D data and some of the challenges to the use of MIDs;
- to explain why case-mix adjustment of EQ-5D data is important in some contexts, and how that might be done; and
- to provide an overview of the use of mapping techniques to link other Patient Reported Outcome (PRO) data to the EQ-5D and EQ-5D values.

### 5.1 Analysing Changes and Differences in EQ-5D Values and EQ VAS Scores

In the analyses described in Chaps. 3 and 4, the objects of interest are EQ-5D values or EQ VAS scores measured at one or more points in time for one person or a group of people. These can be compared between the same person at different time points, which we will call ‘changes’, or between different people or groups of people, which we will call ‘differences.’ When the object of interest is the change or difference itself, analysts should be cautious in the way that they deal with them.

#### 5.1.1 *Defining the Outcome of an Intervention Study as a Change*

In clinical studies of the impact of a health care intervention on health-related quality of life (HRQoL), it is possible to define the outcome in two different ways—the

final state of health or the change from initial to final state. In many contexts, the magnitude as well as direction of the change is the object of interest, but there are some well-known issues about estimating the size of changes directly. These relate to all outcome measures, not just health status or HRQoL (for example Vickers and Altman 2001; Bland and Altman 2011), and to Patient Reported Outcome Measures (PROMs) other than EQ-5D instruments, but the characteristics of EQ-5D and EQ VAS values data mean that they are particularly vulnerable to misleading analysis through misspecification of the underlying analytical model.

The key issue is the relationship between the size of initial, or baseline, health state values and the size of the change in them. The most obvious null hypothesis is that baseline and final mean HRQoL scores are the same, equivalent to a mean change score of zero. However, for conditions where underlying health is deteriorating or the condition is self-limiting, this null hypothesis may not be the correct choice. The size of the change may also be related to the baseline in different ways, depending on both the condition and the treatment. For example, if the treatment leads to the same final health state for all patients, the change will be greater, the lower the initial health state; if the treatment is less successful for those with a poorer health state, the change will be greater, the higher the initial health state. Only if the change is constant whatever the initial health state will there be no such complicating relationship.

Figure 5.1 illustrates this point. The horizontal and vertical axes show baseline and final EQ 5D values scores respectively. The solid 45° line shows points where baseline and final are the same, which would be the null hypothesis for patients whose condition is neither improving or deteriorating. The dotted line shows a different

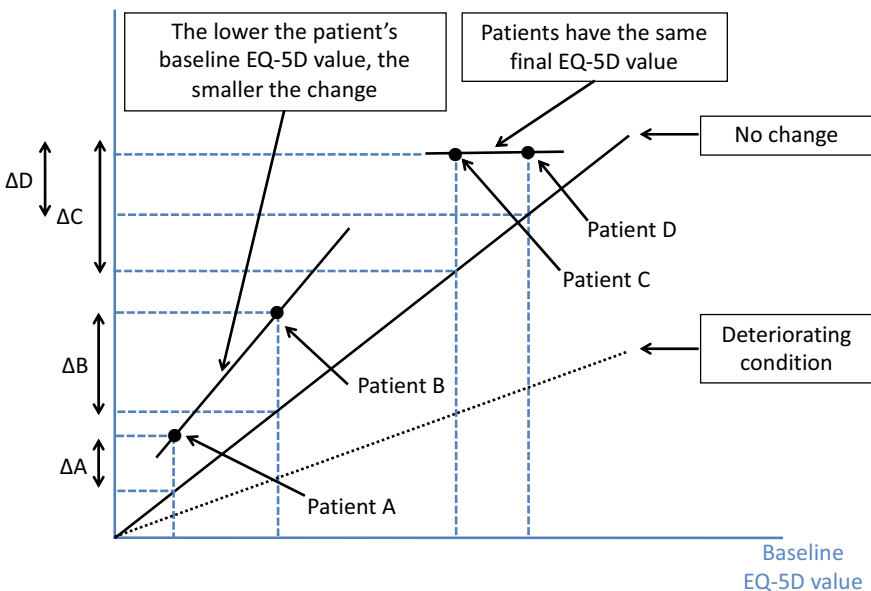


Fig. 5.1 Stylised example of treatment effects

assumption: that the condition would result in a deterioration of the patients' health over time if untreated. The two solid lines above the 'No Change' line show different relationships between baseline and final scores for two different treatments. Patients A and B undergo a treatment for which the outcomes are better for patients whose baseline score is higher. Assuming the null hypothesis of no improvement or deterioration without treatment, the change after treatment for patient A, who has a lower baseline score than patient B, is smaller ( $\Delta A$ ) than that for patient B ( $\Delta B$ ). Patients C and D undergo a treatment which results in the same final score for all patients. Again, assuming that there would be no change without treatment, the change for patient C, who has a lower baseline score than patient D, is bigger ( $\Delta C$ ) than that for patient D ( $\Delta D$ ).

The special problem that this raises for both EQ-5D values and EQ VAS scores is that the existence of fixed end-points—0 and 100 for the EQ VAS; 1 and the value of the worst health state for EQ-5D values—places limits on the possible size of change. (The same is true for any outcome measure that has the same properties.) For EQ-5D values, there is an additional problem that the distribution of scores at both baseline and final assessment may not be smooth because of the discrete nature of the EQ-5D health states from which the scores are calculated. These two issues also mean that there may not be the necessary linear relationship between the baseline and final outcome scores that would permit calculation of a single change-based effect size.

The recommendations are therefore to specify carefully the counterfactual to the observed change or difference, or where possible to ensure that there are control groups from which this can be directly measured, and to ensure that appropriate methods are used to transform the distribution of EQ-5D values into a form amenable to statistical analysis.

### 5.1.2 *Minimal Important Differences (MIDs)*

The calculation of Minimal Important Differences (MIDs) for HRQoL or PRO measures, including the EQ-5D, is a topic on which there is currently no consensus, either to its usefulness or the best methods for its estimation. Those who wish to use or estimate MIDs are therefore advised to consult two review articles, one on MIDs in general (King 2011) and the other specifically on the EQ-5D (Coretti et al. 2014). Here we summarise some of the issues.

The term MID is used here, but other terms are used which, as King points out, may differ slightly in their definitions and meaning such as *minimal clinically important difference (MCID)*, *clinically important difference*, *minimally detectable difference*, *minimum detectable change*, and *subjectively significant difference*. The most widely quoted definition of the concept is of a MCID (Jaeschke et al. 1989), but an updated MID-specific version of this (Guyatt et al. 2002) is “the smallest difference in score in the domain of interest that patients perceive as important, either beneficial or harmful, and which would lead the clinician to consider a change in the patient's management”. Coretti et al. make use of a different term, *the smallest worthwhile*

*effect* (SWE), defined by Ferreira et al. (2012) as “the smallest beneficial effect justifying costs, risks and inconveniences of an intervention.”

There are three key questions to address when deciding whether and how to use MIDs with an HRQoL or PRO measure such as the EQ-5D: What is the purpose of using a MID? What definition should be used for that purpose? and How should the MID be estimated to meet that definition? Although in principle it would be possible to ask these questions about EQ-5D health states, in practice they have only been explored for EQ-5D values and to a lesser extent to EQ VAS scores, so this guide has the same limitation.

In answering these questions, it is essential to note that EQ-5D values have a feature that distinguishes them from some other measures. They already embody a measure of importance as perceived by a group of people, usually a general population, based on their preferences for different health states (see Chap. 4). The values are estimated from an underlying continuous value function at discrete points on the value scale identified by the EQ-5D health states. Any differences in the underlying values, however small, are therefore important in that they indicate a difference that would be preferred or non-preferred by the person affected, other things being equal. Similar arguments apply to the scores generated by the EQ VAS.

A wider definition of importance, such as whether a change is worthwhile given the perceived importance to patients and resource costs of making the change and the duration for which the change is experienced, requires information that is not contained within the EQ-5D values or EQ VAS scores themselves. This suggests that there is no conceptual basis for a MID for EQ-5D values or EQ VAS scores in terms of *desirability*; however, it may be possible to base a MID on whether in practice differences and changes in the EQ-5D values or EQ VAS score are *detectable*. As King points out, this concept of ‘minimally detectable’ differences or changes has two separate bases. One is psychometric, and concerns whether a difference is capable of being perceived by people. The other is statistical, concerning measurement error, the precision with which perceived differences are recorded.

## Using EQ-5D MIDs

### *Using EQ-5D MIDs for decision making with individual patients*

As noted, the basis for an EQ-5D MID to judge the importance, in terms of desirability, of differences between or changes in health states is weak. A further problem for using this with individual patients is that they may not share the preferences of the average patient or member of the general population about their health. With respect to detectability, the ability to observe changes or differences in EQ-5D values is entirely based on detection of changes to the EQ-5D health states, and the calculation of a summary index in the form of an EQ-5D value may obscure rather than illuminate the nature of the change.

### *Using EQ-5D MIDs for decision making about populations*

Again, it is not possible to judge, in terms of desirability, whether an observed difference or change in EQ-5D values or EQ VAS scores is important without further

information. With respect to detectability, there is also a problem arising from how observations for individual people are aggregated to give a population score, exacerbated in the case of the EQ-5D values by their discrete nature. A population average MID will depend both on the size of changes to each individual person's health state and the number of people experiencing different levels of change. As an extreme example, if all but one member of a group recorded a change of EQ-5D values at the MID value and the exception scored below that, the population would be judged as having a difference below the MID. Comparing the mean to the MID would give a misleading account of the clinical importance of the overall observed differences.

#### *Using EQ-5D MIDs for clinical research*

A proposed use of MIDs is in determining the most efficient sample size for a clinical trial, based on the desired probabilities of avoiding type 1 and type 2 errors. The aim is to ensure that trials are not over-powered, and generate statistically significant differences that have no clinical significance. A trial powered to detect differences at the level of the MID would be the correct approach for a trial for which HRQoL was the primary endpoint and was the sole determinant of clinical decision making. However, the MID is less useful for trials that have a different primary endpoint or where clinical decision making is not independent of factors other than a difference in HRQoL. In addition, it is again necessary to take account of the distribution of observed differences in EQ-5D values, as using an individually-based MID may be misleading about the total benefit over all patients.

#### **MID estimation methods**

A common finding of the different methods described below is that there is no identifiable single MID for EQ-5D values or EQ VAS scores. Instead, estimates differ by population, patient group, clinical context and sociodemographic factors; and might vary depending on whether health is improving or worsening. It is possible to calculate a score which is an average over different patient populations, such as the widely-quoted estimate by Walters and Brazier (2005) for EQ-5D values (which is described below), but although this is an interesting statistic, the size of the variability between different estimates means that an average EQ-5D MID should not be used for any of the purposes described above.

#### *Patient rating of change*

The most common and direct method of meeting the aim of assessing patients' own perceptions of the importance of differences in their health is to quiz them specifically about that, using a *global transition question*. This is a retrospective assessment by patients of the change in their health between two points, at each of which their current health has been assessed using the HRQoL or PRO instrument. For example, Walters and Brazier (2005) re-analysed 11 studies in different clinical areas that collected both EQ-5D and SF-36 data at different time points. They compared the differences between EQ-5D values with a question taken from the SF-36, asking if their general health was much better, somewhat better, stayed the same, somewhat worse or much worse, compared to the last time they were assessed. Those who answered somewhat

better or somewhat worse were considered as having experienced a change equivalent to the MID.

This method relies on the global transition question identifying the minimum perception that patients can have, which is in reality determined by the fixed wording of the text of the permitted answers. For example, patients are likely to have different thresholds for deciding that they have any improvement or deterioration at all, and also different perceptions of the boundaries between ‘somewhat’ and ‘much’. If these do not match the boundaries between the descriptions contained in the EQ-5D health states, then the calculated EQ-5D value changes for the ‘somewhat’ categories may not reflect the true size of the minimum differences that patients perceive. In addition, global transition questions are affected by the ability of patients to recall their previous health state accurately and may be more subject to acquiescence bias and response shift (Sprangers and Schwartz 1999; Kamper et al. 2009).

### *Clinical anchors*

Another common method of defining a MID is to examine the scores of patients classified according to a different measure of their clinical status. The rationale is that for clinical decision making, clinicians may have more confidence with an HRQoL measure if it is related to more familiar, clinically-focussed and well-validated measures. For example, Pickard et al. (2007) calculated the mean EQ-5D values and EQ VAS scores for cancer patients in the different grades of two clinical measures, the Eastern Cancer Oncology Group (ECOG) and the Functional Assessment of Cancer Therapy General (FACT-G). The differences between the mean scores between different grades, ordered according to severity of the condition, provides MID estimates as a range and average.

This method emphasises the clinical decision-making aspect of the definition of a MID rather than the idea that it should reflect patients’ own perception of the importance of change. It therefore depends on an assumption that the clinical anchor measure correctly distinguishes between important and unimportant changes in health states.

### *Distribution-based*

Some estimates of the MID are based on statistics that describe the distribution of health states in a patient population, in particular the standard error of measurement (SEM) and the effect size (ES). Pickard et al. (2007) also estimated MIDs for EQ-5D values and EQ VAS scores using both of these approaches, stratified again according to FACT-G and ECOG grades. The SEM is based on reliability of the HRQoL or PRO instrument, usually measured with respect to test-retest reliability, the distribution around a true score of repeated assessments assuming no memory effect or other contextual changes, which is regarded as a fixed psychometric property of the instrument. An alternative measure is reliability based on internal consistency measured by Cronbach’s alpha, which is what Pickard et al. used because of the scarcity of test-retest information for the EQ-5D.

The ES is calculated as the mean difference in HRQoL divided by the between-person standard deviation. Pickard et al. based their MIDs on the criterion of one-half of the standard deviation (SD), although one-third and one-fifth SD are also commonly used (King 2011).

These methods again do not reflect patients' perceptions of importance, and unless they are stratified in the way used by Pickard et al. also do not reflect importance as defined by a clinician for use in decision-making.

#### *Instrument-defined*

Luo et al. (2010) and McClure et al. (2017) have calculated MIDs for the 3L and 5L respectively using a method that does not require empirical EQ-5D or other data. It is calculated, for a specified value set, as the smallest difference in the values of any pair of health states, over all possible pairs. It is therefore the smallest possible observed difference in values either for a person whose EQ-5D health state is captured at two different times or for two people at the same time.

This highlights an important property of a value set, and is useful in examining the comparative performance of different value sets. However, it does not match with the usual definitions of a MID and it is not obvious how it might be used for any of the purposes described above. The differences between the values of different health states are entirely determined by perceived differences in the descriptions that the health states are given. This MID therefore does not reflect the smallest score that people find important, but the smallest difference between the health state descriptions, which is fixed by the descriptive system itself, not by the people who value them. As importantly, it is based on an assessment of health state differences for individual people, and in a group or population context, it is highly vulnerable to the problem outlined above of the mean giving a misleading account of overall clinical importance.

The overall recommendation for MIDs is that the purpose of using a MID in a particular context should be carefully considered, that a precise definition for the MID is derived from that purpose, and that the methods used to estimate that MID fit with the definition adopted.

### ***5.1.3 Case-Mix and Risk Adjustment of EQ-5D Data***

Although we refer here to case-mix adjustment, the principles also apply to the related concept of risk adjustment. Adjusting HRQoL or PRO scores for the differing characteristics of patients and external factors is often essential in making comparisons of outcomes. For example, when comparing the average observed EQ-5D value or EQ VAS score changes after treatment for patients in different hospitals, it is important to account for factors that affect outcomes but are not due to variations in the quality of care. One such factor may be the average age of patients treated, which may differ between different providers and affect the outcomes that can be achieved. To obtain

a fair comparison of the outcomes of different hospitals, they should be adjusted to take account of the mix of cases that the hospitals see.

There are many different methods for calculating case-mix adjustments, including stratification and direct and indirect standardisation. Stratification refers to calculation of outcomes for subgroups of a population defined according to key characteristics that might affect outcomes, such as age, sex and ethnicity. Direct standardisation calculates outcomes for different units, such as hospitals, adjusted by comparison of the levels of the case-mix variables to those in a known reference population. Indirect standardisation uses, instead of a known reference population, the average level of the variables for the units as a whole. Here we give an example of how the United Kingdom's National Health Service (NHS) adjusts for case-mix in its PROMs programme (Nuttall et al. 2015; Department of Health 2012; NHS England Analytical Team 2013) using the indirect standardisation method.

The NHS case-mix adjustment method has two stages. The average impact of case-mix variables on EQ-5D values or EQ VAS scores is calculated over all patients using regression analysis. The regression coefficients are used to estimate, for each health care provider, the average EQ-5D value or EQ VAS score that would be expected for its mix of those variables. From this, the difference between expected and actual outcomes is calculated for each provider.

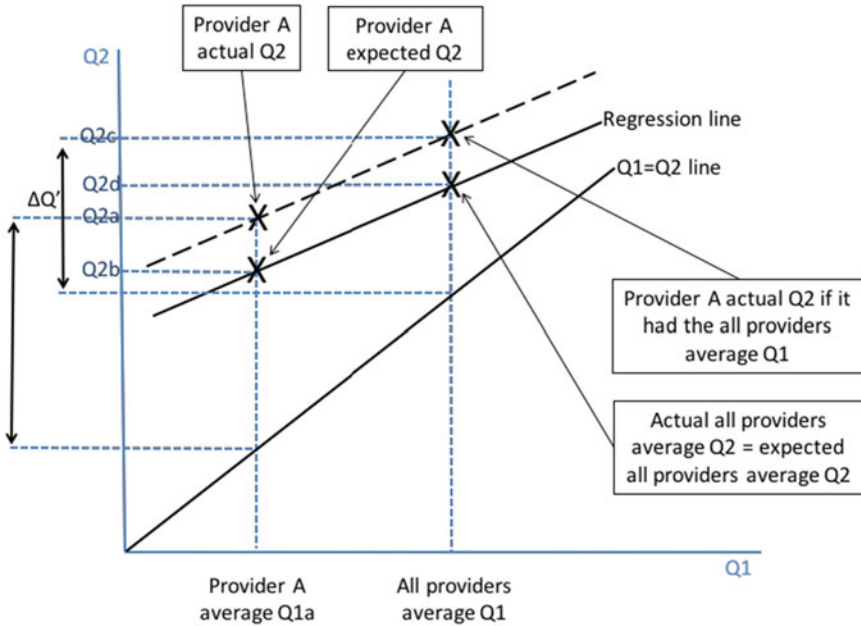
This is regarded as a measure of a provider's performance, but the 'expected' outcome is for a hypothetical provider that has the same case-mix, and does not compare the provider with other real providers. Each provider's outcomes are therefore transformed so that they can be compared to a standard case-mix, which is the mean level of the case-mix variables over all providers. This also generates the all-providers average EQ-5D value or EQ VAS score, by definition.

Figure 5.2 illustrates this, using a very simple case-mix adjustment to the post-treatment EQ-5D value or EQ VAS score (Q2), taking account of the pre-treatment value of the score (Q1). An observation on the  $Q1 = Q2$  line would mean that there had been no change in the average EQ-5D value or EQ VAS score. The hypothetical regression line lies above this, meaning that at all levels of Q1 there is on average an improvement following surgery. Q2 is higher with higher Q1, but the size of the improvement (the difference between Q2 and Q1) is smaller with higher Q1.

For provider A, its average post-surgery EQ-5D value or EQ VAS score is Q2a, so that the change in the EQ-5D value or EQ VAS score unadjusted for case-mix is  $\Delta Q = Q2a - Q1a$ . Its expected EQ-5D value or EQ VAS score is Q2b and it therefore has performed better than would be expected for a provider that had the same case-mix.

Performance can be quantified as  $Q2a - Q2b$ ; if this is positive, the provider achieves on average results greater than those predicted; negative if worse than predicted; and zero if as predicted. This difference is applied to the all-provider Q2 EQ-5D value or EQ VAS score, which is Q2d, to give the estimated actual EQ-5D value or EQ VAS score for Provider A if it had the all-provider case-mix. This EQ-5D value or EQ VAS score, Q2c, is calculated so that  $Q2c - Q2d = Q2a - Q2b$ , which means  $Q2c = Q2d + (Q2a - Q2b)$ . The relevant Q1 comparator for this is the all-provider Q1 EQ-5D value or EQ VAS score, so the case-mix adjusted change in the EQ-5D value or EQ VAS score for Provider A is  $\Delta Q' = Q2c - Q1$ .





**Fig. 5.2** Stylised example of case-mix adjustment. This figure is taken from Chapter 3 of Appleby et al. (2015)

Amongst the problems with this method are those outlined in Sect. 5.1.1 concerning the assumed counterfactual to the observed changes and the effect of fixed end-points and discrete EQ-5D values on the distribution of changes and their relationship with the pre-surgery EQ-5D values or EQ VAS scores.

Case-mix adjustments can change the estimated outcomes for different units such that a very different assessment is made of their relative performance. For example, Appleby et al. (2015) showed that using a case-mix adjustment for changes in EQ-5D values in the English NHS PROMs programme reduced the range of average hospital scores and the size of their variability around the national average. More importantly, it produces a different performance ranking of hospitals in terms of health gain, as individual hospitals’ adjusted and unadjusted gains differ considerably in many cases.

## 5.2 Mapping

In this context, mapping refers to methods that are used to convert the responses of one HRQoL or PRO measure to those of a different measure. The most usual application for the EQ-5D is based on an interpretation of EQ-5D values as numbers representing the values that people attach to health states, which have cardinal measurement

properties such that they can be used to calculate Quality Adjusted Life Years, which can be used as the denominator in an Incremental Cost Effectiveness Ratio. Mapping is used to convert data from a measure that does not have these properties, such as a condition specific instrument, to EQ-5D values. This takes the form of an algorithm which is applied to the source measure and generates EQ-5D values. Mapping could also be used simply to translate the responses given in another HRQOL measure into EQ-5D health states.

Mapping is also used to convert between the values of the 3L and 5L versions. However, we will not discuss the methods used for this as they concern valuation of health states, which is outside of the scope of this review. At the time of writing, there are no definitive guidelines for those who wish to convert 3L to 5L or vice versa, and a continuing debate about the best methods. Those wishing to make use of such mapping are advised to consult the most up-to-date literature; current key references include van Hout et al. (2012), Hernandez-Alava et al. (2017) and Dakin et al. (2018).

There are useful statements of good practice in mapping to health state measures that have the value-based and cardinality properties described above from measures that do not (Wailoo et al. 2017), and for reporting those studies (Petrou et al. 2015). There is also an online database of existing mapping studies (Dakin 2013). Those who wish to undertake mapping or use existing mapping algorithms are advised to consult those papers, and here we simply summarise some of the issues. It should be emphasised that mapping is a second-best approach that produces only an approximation to true EQ-5D values. The availability of a mapping algorithm for a particular measure can never be a justification for failing to collect EQ-5D health state data as well as or instead of that measure.

The earliest studies that undertook mapping were often based on direct judgements by clinical experts, patients or researchers about the correspondence between the descriptive systems of the source measure and EQ-5D values. This is not now regarded as good practice. Acceptable mapping methods require data that have been collected from respondents who have completed both the source measure and the EQ-5D.

There is a broad division of mapping methods between those that map directly to EQ-5D values and those, known as *response mapping*, that map to EQ-5D health states, from which EQ-5D values are calculated using a value set. For the direct method, it is possible simply to assign EQ-5D values for the health state recorded by each respondent to the category or score that they report for the source measure, and calculate the mean over all respondents. However, this method is restrictive, because it only enables mapping for those health states present in the sample in large enough numbers. It is also known that other patient characteristics and health and treatment condition variables may impact on the mapping. As a result, it is regarded as best practice to use a regression-based method to ensure that the mapping algorithm is both more comprehensive and more precise.

The response mapping method has the advantage, when *using* a mapping algorithm, that it produces an algorithm that generates EQ-5D health states, to which any value set can be applied, while the direct method is specific to a particular value set.

The direct method has the advantage, when *generating* a mapping algorithm, that in estimating the relationship between the source measure and the EQ-5D, the response or dependent variable—EQ-5D values—can be treated as a continuous variable. The response mapping method is based on categories—EQ-5D health states—that do not even have ordinal properties. This is a problem because it potentially requires a data set large enough to contain a meaningfully-large observation for each of the 243 (3L) or 3125 (5L) health states. However, in practice this problem is dealt with by assuming that the level recorded in each dimension is independent of the level recorded in other dimensions. This permits estimation of five separate ordered dependent variables, which is statistically much more amenable to analysis.

## References

- Appleby J, Devlin N, Parkin D (2015) Using patient reported outcomes to improve health care. Wiley Blackwell
- Bland JM, Altman DG (2011) Comparisons against baseline within randomised groups are often used and can be highly misleading. *Trials* 12(1):264
- Coretti S, Ruggeri M, McNamee P (2014) The minimum clinically important difference for EQ-5D index: a critical review. *Expert Rev Pharmacoecon Outcomes Res* 14(2):221–233
- Dakin H (2013) Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. *Health Qual Life Outcomes* 11:151
- Dakin H, Abel L, Burns R, Yang Y (2018) Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. *Health Qual Life Outcomes* 16(1):31
- Department of Health (2012) Patient Reported Outcome Measures (PROMs) in England: The case-mix adjustment methodology. London: Department of Health
- Ferreira ML, Herbert RD, Ferreira PH et al (2012) A critical review of methods used to determine the smallest worthwhile effect of interventions for low back pain. *J Clin Epidemiol* 65(3):253–61
- Guyatt GH, Osaba D, Wu AW et al (2002) Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 77(4):371–383
- Hernandez Alava M, Wailoo A, Pudney S (2017) Methods for mapping between the EQ-5D-5L and the 3L for technology appraisal: report by the NICE Decision Support Unit. School of Health and Related Research, University of Sheffield, UK, Health Economics and Decision Science
- van Hout B, Janssen MF, Feng Y et al (2012) Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 15:708–15
- Jaeschke R, Singer J, Guyatt GH (1989) Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 10(4):407–15
- Kamper SJ, Maher CG, Mackay G (2009) Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther* 17(3):163–70
- King MT (2011) A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 11(2):171–184
- Luo N, Johnson JA, Coons SJ (2010) Using instrument-defined health state transitions to estimate minimally important differences for four preference-based health-related quality of life instruments. *Med Care* 48:365–71
- McClure NS, Al Sayah F, Xie F, Luo N, Johnson JA (2017) Instrument-defined estimates of the minimally important difference for EQ-5D-5L index scores. *Value Health* 20:644–650
- NHS England Analytical Team (2013) Patient Reported Outcome Measures (PROMs): An alternative aggregation methodology for case-mix adjustment. NHS England. <http://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-agg-meth-adju.pdf>

- Nuttall D, Parkin D, Devlin N (2015) Inter-provider comparison of patient-reported outcomes: developing an adjustment to account for differences in patient case mix. *Health Econ* 24(1):41–54
- Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R, Gray A (2015) The MAPS reporting statement for studies mapping onto generic preference-based outcome measures: explanation and elaboration. *Pharmacoeconomics* 33:993–1011
- Pickard AS, Neary MP, Cella D (2007) Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health Qual Life Outcomes* 5:70
- Sprangers MA, Schwartz CE (1999) Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med* 48:1507–1515
- Vickers AJ, Altman DG (2001) Analysing controlled trials with baseline and follow up measurements. *BMJ* 323(7321):1123–1124
- Wailoo AJ, Hernandez-Alava M, Manca A, Mejia A, Ray J, Crawford B, Botteman M, Busschbach J (2017) Mapping to estimate health-state utility from non-preference-based outcome measures: an ISPOR good practices for outcomes research task force report. *Value Health* 20:18–27
- Walters SJ, Brazier JE (2005) Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 14:32

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

