

Commentary

Open Access

Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change

Henrica C de Vet*, Caroline B Terwee, Raymond W Ostelo, Heleen Beckerman, Dirk L Knol and Lex M Bouter

Address: EMGO Institute, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands

Email: Henrica C de Vet* - hcw.devet@vumc.nl; Caroline B Terwee - cb.terwee@vumc.nl; Raymond W Ostelo - r.ostelo@vumc.nl; Heleen Beckerman - h.beckerman@vumc.nl; Dirk L Knol - d.knol@vumc.nl; Lex M Bouter - lm.bouter@vumc.nl

* Corresponding author

Published: 22 August 2006

Received: 26 July 2006

Health and Quality of Life Outcomes 2006, **4**:54 doi:10.1186/1477-7525-4-54

Accepted: 22 August 2006

This article is available from: <http://www.hqlo.com/content/4/1/54>

© 2006 de Vet et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Changes in scores on health status questionnaires are difficult to interpret. Several methods to determine minimally important changes (MICs) have been proposed which can broadly be divided in distribution-based and anchor-based methods. Comparisons of these methods have led to insight into essential differences between these approaches. Some authors have tried to come to a uniform measure for the MIC, such as 0.5 standard deviation and the value of one standard error of measurement (SEM). Others have emphasized the diversity of MIC values, depending on the type of anchor, the definition of minimal importance on the anchor, and characteristics of the disease under study. A closer look makes clear that some distribution-based methods have been merely focused on minimally detectable changes. For assessing minimally important changes, anchor-based methods are preferred, as they include a definition of what is minimally important. Acknowledging the distinction between minimally detectable and minimally important changes is useful, not only to avoid confusion among MIC methods, but also to gain information on two important benchmarks on the scale of a health status measurement instrument. Appreciating the distinction, it becomes possible to judge whether the minimally detectable change of a measurement instrument is sufficiently small to detect minimally important changes.

Introduction

Health status questionnaires are increasingly used in medical research and clinical practice. They are attractive because they provide a self-report of patients' perceived health status. However, the meaning of the (changes in) scores on these questionnaires is not intuitively apparent. The interpretation of (change)scores has been a topic of research for almost two decades [1,2]. It is recognized that the statistical significance of a treatment effect, because of its partial dependency on sample size, does not always

correspond to the clinical relevance of the effect. Statistically significant effects are those that occur beyond some level of chance. In contrast, clinical relevance refers to the benefits derived from that treatment, its impact upon the patient, and its implications for clinical management of the patient [2,3]. As a yardstick for clinical relevance one is interested in the minimally important change (MIC) of health status questionnaires. Changes in scores exceeding the MIC are clinically relevant by definition.

Different methods to determine the MIC on the scale of a measurement instrument have been proposed. These methods have been summarized by Lydick and Epstein [4], and recently more extensively by Crosby et al. [5]. Both overviews distinguish distribution-based and anchor-based methods [4,5].

Distribution-based approaches are based on statistical characteristics of the sample at issue. Most distribution-based methods express the observed change in a standardized metric. Examples are the effect size (ES) and the standardized response mean (SRM), where the numerators of both parameters represent the mean change and the denominators are the standard deviation at baseline and the standard deviation of change for the sample at issue, respectively. Another distribution-based measure is the standard error of measurement (SEM), which links the reliability of the measurement instrument to the standard deviation of the population [5]. ES and SRM are relative representations of change (without units), whereas the SEM provides a number in the same units as the original measurement. The major disadvantage of all distribution-based methods is that they do not, in themselves, provide a good indication of the importance of the observed change.

Anchor-based methods assess which changes on the measurement instrument correspond with a minimal important change defined on the anchor [4], i.e. an external criterion is used to operationalize a relevant or an important change. The advantage is that the concept of 'minimal importance' is explicitly defined and incorporated in these methods. A limitation of anchor-based approaches is that they do not, in themselves, take measurement precision into account [4,5]. Thus, there is no information on whether an important change according to an anchor-based method, lies within the measurement error of the health status measurement.

An often used anchor-based method is the one proposed by Jaeschke et al. [2], which defined MIC as the mean change in scores of patients categorized by the anchor as having experienced minimally important improvement or minimally important deterioration. Another anchor-based method, proposed by Deyo and Centor [6], is based on diagnostic test methodology. In this method, the change score on the measurement instrument is considered the diagnostic test and the anchor, dividing the population in persons who are minimally importantly changed and those who are not, is considered the gold standard. At different cut-off values of change scores the sensitivity and specificity are calculated and the MIC value is set at the change value on the measurement, where the sum of the percentages of false positives and false negatives is minimal.

A number of studies have compared anchor-based and distribution-based approaches. Comparisons of these approaches sometimes led to surprisingly similar results. However, in other situations different results were found. The focus of this paper is on explanation of the differences between distribution-based and anchor-based approaches. We will provide arguments for the distinction between minimally detectable change and minimally important change. Appreciating and acknowledging this distinction enhances the interpretation of change scores of a measurement instrument.

Comparison of SEM with anchor-based approaches

A number of studies have compared the value of the SEM with the MIC value derived by an anchor-based approach. A SEM value is easy to calculate, based on the standard deviation (SD) of the sample and the reliability of the measurement instrument: in formula: $SEM = SD \sqrt{1-R}$. As reliability parameter, test-retest reliability or Cronbach's α can be used. In the latter case, SEM can be calculated based on one measurement and it purely represents the variability of the instrument [7]. Test-retest reliability requires two measurements in a stable population. It represents the temporal stability and is therefore more appropriate than Cronbach's α to use in the context of changes in health status which are based on measurements at two different time points [8]. In classical test theory, SEM has a rather stable value in different populations [6].

Several authors showed that a MIC based on patient's global rating as anchor was close to the value of one SEM [9,11]. Cella et al. [12] also observed similar values for SEM and MIC, using clinical parameters as anchor instead of patients' global rating of change. However, Crosby et al. [13] showed that only for patients with moderate baseline values the anchor-based MIC value more or less equalled the SEM value (with adjustment for regression to the mean). With higher baseline values the MIC became considerably larger than one SEM, while with lower baseline values the MIC became much smaller than one SEM. A recent study [14] compared SEM with anchor-based estimations of minimally important change using cross-sectional and longitudinal anchors. No substantial differences were found between these methods, but it should be noted that they only presented anchor-based values when effect sizes were between 0.2 and 0.5 [14].

Wyrwich [15], compared SEM to MIC values determined by an anchor-based approach in two sets of studies which differed on several points. Set A consisted of studies on musculoskeletal disorders like low back pain [16], neck pain [17] and lower extremity disorders [18], while set B included studies on chronic disorders like chronic respiratory disease [10], chronic heart failure [11], and asthma [19]. In addition, set A studies used the ROC method and

studies in set B applied mean change as anchor-based method. And they differed with regard to the definition of 'minimal important change' on the anchor. For set A, the MIC corresponded to 2.3 or 2.6 * SEM, for set B, the MIC values were close to 1*SEM.

In summary, it seems that the proposition that one SEM equals the MIC is not a universal truth.

MIC is a variable concept

1. MIC depends on the definition of 'important change' on the anchor
A patient's self report global rating scale of perceived change has often been used as anchor. Studies determining MIC have used different definitions of 'minimally important' using this anchor. Wyrwich et al. [10,11,19] defined a slight change on the anchor as 'minimally important', consisting of the categories "a little worse/better" and "somewhat worse/better". In their earlier studies Wyrwich et al. even included the category "almost the same, hardly any worse/better" [10,11]. Other authors have defined 'minimal importance' as a larger change on the anchor. Binkley et al. [18] chose the category "moderate improvement" as minimally important. Stratford et al. [17] chose to lay the cut-off point for MIC between "moderate" and "a good deal" of improvement. Others [20-24] have laid the cut-off point for MIC between "slightly improved" and "much improved" on the patient global rating scale. In studies requiring moderate or much improvement, the MIC corresponds to about 2.5 times the SEM value. The differences in set A and B in Wyrwich's study [15] may be partly explained by a different definition of important change on the anchor: set A consists of studies which defined MIC as a good deal better [16] and studies in set B [10,11,19] defined MIC as a little and somewhat better according to the anchor.

The MIC value depends to a great degree on the anchor's definition of *minimal importance*. So, the crucial question, then, is "what is a minimally important improvement or deterioration?" Some authors tend to emphasize *minimal*, while others stress *important* [25]. Remarkably, the reference standard is usually based on the amount of change and little research has focused on the "importance" of the change.

2. MIC depends on the type of anchor

Clinicians may have other opinions about what is important than patients. Therefore, clinician-based anchors may lead to different MIC values. Kosinski et al. [26] used five different anchors to estimate the minimally important differences for the SF-36 in a clinical trial of people with rheumatoid arthritis, and found different MIC values dependent on the anchor used. Some authors [16,20-24] have asked patients' global rating of perceived change in overall health, while others asked to rate the perceived

change separately for each dimension of their measurement instrument [10,19]. For example, in a study determining the MIC of the Chronic Respiratory Disease Scale the patients' global rating has been asked separately for the subscales dyspnoea, fatigue and emotional function [10]. In the rating of change in overall health status patients have to weigh the relative contribution of the different dimensions on their health status. For example, if patients with asthma judge dyspnoea to be much more important for their quality of life than emotional functioning, a small change in dyspnoea will affect the global rating of overall health, while for emotional functioning the change must be larger to be influential. The observed MIC value will be smallest for the anchor that shows the highest correlation with the health status scale under study.

3. MIC depends on baseline values and direction of change

Several studies have shown that the MIC value of a measurement instrument depends on the baseline score on that instrument. This was clearly shown by Crosby et al. [13] who compared the SEM, corrected for regression to the mean, with the anchor-based MIC for various baseline scores of obesity-specific health related quality of life. With higher baseline values MIC became considerably larger than one SEM. Other authors [16,24,27,28] showed that the values of anchor-based MIC for functional status questionnaires in patients with low back pain were dependent on baseline values. Patients with a high level of functional disability at baseline must change more points on the Roland Disability Questionnaire than patients with less functional disability at baseline to consider it an important change. In addition, Van der Roer et al. [24] reported different MIC values for acute and chronic low back pain patients.

Furthermore, there has been discussion whether the MIC for improvement is the same as for deterioration [5]. In some studies the same MIC is reported for patients who improve and patients who deteriorate [2,29,30], but others found different MIC values for improvement and deterioration. Cella et al. [31] demonstrated that cancer patients who reported global worsening had considerably larger change scores on the Functional Assessment of Cancer Therapy (FACT) scale than those reporting comparable global improvements. Also Ware et al. observed that a larger change on the SF-36 was needed for patients to feel worsened than to feel improved [32].

Thus, the MIC is dependent on, among other things, the type of anchor, the definition of 'minimal importance' on the anchor, and on the baseline score which might be an indicator of severity of the disease. Therefore, various authors have suggested to present a range of MIC values [24,26,33-35], to account for this diversity. Hays et al. rec-

commend to use different anchors and to give reasonable bounds around the MIC, rather than forcing the MIC to be a fixed value [33,34].

Distinction between minimally detectable and minimally important changes

Some authors have searched for uniform measures for minimally important changes. Wyrwich and others [10,11] have evaluated whether the one-SEM criterion can be applied as a proxy for MIC. Norman et al. [36] made a systematic review of 38 studies (including 62 effect sizes), and observed, with only a few exceptions, that the MICs for health related quality of life instruments were close to half a standard deviation (SD). This held for generic and disease specific measures and was not dependent on the number of response options.

Norman et al. [36] explain their finding of 0.5 SD by referring to psychophysiological evidence that the limit of people's ability to discriminate is approximately 1 part in 7, which is very close to half a SD. Thus, this criterion of 0.5 SD may be considered a threshold of detection and corresponds more to *minimally detectable* change than to *minimally important* change. Also SEM, based on the test-retest reliability in stable persons, is merely a measure of detectable change [37]. Note that, using the formula $SEM = SD \sqrt{1-R}$, 1 SEM equals 0.5 SD, when the reliability of the instrument is 0.75. Thus, 0.5 SD and SEM clearly alert to the concept of minimally detectable changes.

Wyrwich [15] comparing the two sets of studies showed that if the cut-off point for 'minimal importance' on the anchor is laid between "no change" and "slightly changed", i.e. the first category above no change, together with a complaint-specific anchor, the MIC is close to one SEM. But in this case it focuses more on *minimally detectable* change than *minimally important* change. Wyrwich [15] showed a clear dependency between MIC and cut-off value of 'minimal importance' on the anchor of patients' global rating of perceived change.

Salaffi et al. [38] presented the change on a numerical rating scale for pain using two cut-off points on a patient global impression of change scale. In their opinion, a MIC using "slightly better" as cut-off point on the anchor reflected the minimum and lowest degree of improvement that could be detected, while the cut-off point "much better" refers to a clinically important outcome. Note that the choice of anchor and cut-off point is arbitrary and cannot be based on statistical characteristics.

Interpretation and applicability

We believe that the confusion about MIC will decrease if the distinction between minimally detectable and minimally important change is appreciated and acknowledged.

In statistical terms, the minimally detectable change (MDC), also called smallest detectable change or smallest real change [37] shows which changes fall outside the measurement error of the health status measurement (either based on internal or test-retest reliability in stable persons). It is represented by the following formula: $MDC = 1.96 * \sqrt{2} * SEM$, where the 1.96 derives from the 95% confidence interval of no change, and $\sqrt{2}$ is included because two measurements are involved in measuring change [37].

As a different concept, the MIC value depicts changes which are considered to be minimally important by patients, clinicians, or relevant others. The SEM, the minimally detectable change and the minimally important change are all important benchmarks on the scale of the measurement instrument, which helps with the interpretation of change scores.

Appreciating the distinction, we can answer the important question whether a health status measurement instrument is able to detect changes as small as the MIC value. This application is shown in a study on measurement instruments for low back pain [27] and for visual impairments [39].

Conclusion

Some distribution-based methods to assess MIC have been more focussed on minimally detectable changes than on minimally important changes. For assessing minimally important changes, anchor-based methods are preferred, as they include a definition of what is minimally important. Acknowledging the distinction between minimally detectable and minimally important changes is useful, not only to avoid confusion among MIC methods, but also to gain information on two important benchmarks on the scale of a health status measurement instrument. Moreover, it becomes possible to judge whether the minimally detectable change of a measurement instrument is sufficiently small to detect minimally important changes.

References

1. Jacobson N, Follette W, Revenstorf D: **Toward a standard definition of clinically significant change.** *Behavior Therapy* 1986, **17**:308-311.
2. Jaeschke R, Singer J, Guyatt GH: **Measurement of health status. Ascertaining the minimal clinically important difference.** *Control Clin Trials* 1989, **10**:407-415.
3. Jacobson NS, Truax P: **Clinical significance: a statistical approach to defining meaningful change in psychotherapy research.** *J Consult Clin Psychol* 1991, **59**:12-19.
4. Lydick E, Epstein RS: **Interpretation of quality of life changes.** *Qual Life Res* 1993, **2**:221-226.
5. Crosby RD, Kolotkin RL, Williams GR: **Defining clinically meaningful change in health-related quality of life.** *J Clin Epidemiol* 2003, **56**:395-407.
6. Deyo RA, Centor RM: **Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance.** *J Chron Dis* 1986, **39**:897-906.

7. Nunnally JC, Bernstein IH: *Psychometric theory* New York: McGraw-Hill; 1994.
8. Schmidt FL, Le H, Iliis R: **Beyond alpha: an empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs.** *Psychol Methods* 2003, **8**:206-224.
9. Norquist JM, Fitzpatrick R, Jenkinson C: **Health-related quality of life in amyotrophic lateral sclerosis: determining a meaningful deterioration.** *Qual Life Res* 2004, **13**:1409-1414.
10. Wyrwich KW, Tierney WM, Wolinsky FD: **Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life.** *J Clin Epidemiol* 1999, **52**:861-873.
11. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD: **Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life.** *Med Care* 1999, **37**:469-478.
12. Cella D, Eton DT, Fairclough DL, Bonomi P, Heyes AE, Silberman C, Wolf MK, Johnson DH: **What is a clinically meaningful change on the Functional Assessment of Cancer Therapy-Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592.** *J Clin Epidemiol* 2002, **55**:285-295.
13. Crosby RD, Kolotkin RL, Williams GR: **An integrated method to determine meaningful changes in health-related quality of life.** *J Clin Epidemiol* 2004, **57**:1153-1160.
14. Yost KJ, Cella D, Chawla A, Holmgren E, Eton DT, Ayanian JZ, West DW: **Minimally important differences were estimated for the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) instrument using a combination of distribution- and anchor-based approaches.** *J Clin Epidemiol* 2005, **58**:1241-1251.
15. Wyrwich KW: **Minimal important difference thresholds and the standard error of measurement: is there a connection?** *J Biopharm Stat* 2004, **14**:97-110.
16. Stratford PW, Binkley JM, Riddle DL, Guyatt GH: **Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1.** *Phys Ther* 1998, **78**:1186-1196.
17. Stratford PW, Riddle DL, Binkley JM, Spadoni G, Westaway MD, Padfield B: **Using the neck disability index to make decisions concerning individual patients.** *Physiother Can* 1999, **51**:107-112.
18. Binkley JM, Stratford PW, Lott SA, Riddle DL: **The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network.** *Phys Ther* 1999, **79**:371-383.
19. Wyrwich KW, Tierney WM, Wolinsky FD: **Using the standard error of measurement to identify important changes on the Asthma Quality of Life Questionnaire.** *Qual Life Res* 2002, **11**:1-7.
20. Beurskens AJ, de Vet HC, Koke AJ, Lindeman E, van der Heijden GJ, Regtop W, Knipschild PG: **A patient-specific approach for measuring functional status in low back pain.** *J Manipulative Physiol Ther* 1999, **22**:144-148.
21. Davidson M, Keating JL: **A comparison of five low back disability questionnaires: reliability and responsiveness.** *Phys Ther* 2002, **82**:8-24.
22. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL: **Defining the clinically important difference in pain outcome measures.** *Pain* 2000, **88**:287-294.
23. Ostelo RW, de Vet HC, Knol DL, van den Brandt PA: **24-item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery.** *J Clin Epidemiol* 2004, **57**:268-276.
24. Van der Roer N, Ostelo RW, Bekkering GE, van Tulder MW, de Vet HC: **Minimal clinically important change for different outcome measures in patients with non-specific low back pain.** *Spine* 2006, **31**:578-582.
25. Sloan JA, Cella D, Hays RD: **Clinical significance of patient-reported questionnaire data: another step toward consensus.** *J Clin Epidemiol* 2005, **58**:1217-1219.
26. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE Jr: **Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis.** *Arthritis Rheum* 2000, **43**:1478-1487.
27. Hagg O, Fritzell P, Nordwall A: **The clinical importance of changes in outcome scores after treatment for chronic low back pain.** *Eur Spine J* 2003, **12**:12-20.
28. Riddle DL, Stratford PW, Binkley JM: **Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 2.** *Phys Ther* 1998, **78**:1197-1207.
29. Juniper EF, Guyatt GH, Willan A, Griffith LE: **Determining a minimal important change in a disease-specific Quality of Life Questionnaire.** *J Clin Epidemiol* 1994, **47**:81-87.
30. Redelmeier DA, Guyatt GH, Goldstein RS: **Assessing the minimal important difference in symptoms: a comparison of two techniques.** *J Clin Epidemiol* 1996, **49**:1215-1219.
31. Cella D, Hahn EA, Dineen K: **Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening.** *Qual Life Res* 2002, **11**:207-221.
32. Ware JE, Snow K, Kosinski M, Gandek B: *SF-36 Health Survey: Manual and Interpretation Guide* Boston: The Health Institute; 1993.
33. Hays RD, Woolley JM: **The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it?** *Pharmacoeconomics* 2000, **18**:419-423.
34. Hays RD, Farivar SS, Liu H: **Approaches and recommendations for estimating minimally important differences for health-related quality of life measures.** *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2005, **2**:63-67.
35. Ostelo RW, de Vet HC: **Clinically important outcomes in low back pain.** *Best Pract Res Clin Rheumatol* 2005, **19**:593-607.
36. Norman GR, Sloan JA, Wyrwich KW: **Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation.** *Med Care* 2003, **41**:582-592.
37. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek ALM: **Smallest real difference, a link between reproducibility and responsiveness.** *Qual Life Res* 2001, **10**:571-578.
38. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W: **Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale.** *Eur J Pain* 2004, **8**:283-291.
39. de Boer MR, de Vet HC, Terwee CB, Moll AC, Volker-Dieben HJ, van Rens GH: **Changes to the subscales of two vision-related quality of life questionnaires are proposed.** *J Clin Epidemiol* 2005, **58**:1260-1268.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

