



Swansea University
Prifysgol Abertawe



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in :
United European Gastroenterology Journal

Cronfa URL for this paper:
<http://cronfa.swan.ac.uk/Record/cronfa19097>

Paper:

Alrubaiy, L., Hutchings, H. & Williams, J. (2014). Assessing patient reported outcome measures: A practical guide for gastroenterologists. *United European Gastroenterology Journal*

<http://dx.doi.org/10.1177/2050640614558345>

This article is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Authors are personally responsible for adhering to publisher restrictions or conditions. When uploading content they are required to comply with their publisher agreement and the SHERPA RoMEO database to judge whether or not it is copyright safe to add this version of the paper to this repository.

<http://www.swansea.ac.uk/iss/researchsupport/cronfa-support/>

**Assessing Patient Reported Outcome Measures (PROMs): A practical
guide for gastroenterologists**

Laith Alrubaiy*, Hayley Anne Hutchings, John Gordon Williams

LA, Academic clinical specialty trainee in Gastroenterology and Medicine

HAH, Associate Professor in Health Services Research

JGW, Professor of Health Services Research and Consultant
Gastroenterologist

Patient and Population Health and Informatics Research

College of Medicine, Swansea University,

Swansea SA2 8PP, UK

*Corresponding author:

Dr. Laith Alrubaiy

Room 220, Grove Building

College of Medicine Swansea University, Swansea SA2 8PP

Email: l.alrubaiy@swansea.ac.uk

Tel: 0044 7809670683

Keywords : Patients reported outcome measures, PROM, quality of life

Word count : Main text 3084 words. Abstract 167 words

ABSTRACT:

Gastrointestinal illnesses cause physical, emotional, and social impact on patients. Patient Reported Outcome Measures (PROMs) are increasingly used in clinical decision-making, clinical research, and approval of new therapies. In the last decade, there has been a rapid increase in the number of PROMs in gastroenterology and therefore, the choice between which of these PROMs to use can be difficult. Not all PROM instruments currently used in research and clinical practice in gastroenterology have gone through a rigorous development methodology. New drugs and therapies will not have access to the market if the PROMs used in their clinical trials are not validated according to the guidelines of the international agencies. Therefore, it is important to know the required properties of PROMs when choosing or evaluating a drug or a clinical intervention. This paper reviews the current literature on how to assess the validity and reliability of PROMs. It summarizes the required properties into a practical guide for gastroenterologists to use in assessing an instrument for use in clinical practice or research.

INTRODUCTION:

More than a century ago, the first health outcome measure was proposed by Florence Nightingale by classifying patients into relieved, unrelieved and dead (1). Other guides such as mortality rates have historically been used to measure health outcomes in the population (2). However, the definition of health has changed in the past century to include a wider view of outcomes, which includes freedom from disease, ability to perform daily activities, happiness, social and emotional well being, and quality of life. The World Health Organization (WHO) has defined health as “physical, mental, and social well-being, and not merely the absence of disease and infirmity” (3). As a result numerous measures have been developed in an attempt to quantify health. Health outcome measures are tools used to evaluate an individuals’ health using different health related parameters. Patient Reported Outcomes Measures (PROMs) are instruments that are completed by patients and capture one or more aspects of health (4, 5). The use of PROMs formally to monitor surgical outcomes in England has been an important development (6). Since 2007, the Department of Health has required the routine measurement of patient- reported health outcomes for all NHS patients via its PROMs programme (5, 6). PROMs are increasingly used in decision-making to encourage a patient-centered approach (5, 7). For this reason, PROMs that are chosen and used in practice must be valid, reliable, and clinically useful measures.

There are over 100 PROMs applicable to gastroenterological disorders which are described in the Gastrointestinal patients reported outcomes (GI-PRO) database (4). The classical physician based health outcome measures are

now increasingly being replaced by patient-reported outcome measures (PROMs), which enable the assessment and monitoring of disease or treatment effects from the patients' own perception, rather than the health professionals' judgments (which does not always reflects patients' views) (8, 9).

A doctor's ability to interpret and apply PROMs in clinical practice has a great potential to contribute to a better understanding of the patients' well-being. There is an increasing use of PROMS in both research and clinical work in gastroenterology. It is important differentiate between PROMS for functional disorders and those for organic disorders. Patient reported outcomes measures should be used in identification of the symptomatic profiles, diagnosis and treatment of functional disorders such as Post Prandial Distress Syndrome, Epigastric Pain Syndrome, Chronic Idiopathic Nausea, Excessive Belching, IBS and other functional diseases (10). The lack of objective measurable markers of symptoms improvement, such as stool frequency and rectal bleeding, means the evaluation of treatment response has to be based on the patients' reporting of symptoms. The Rome III criteria (11, 12) are very useful to assess the outcomes of new treatments for functional gastrointestinal disorders.

Nowadays, PROMs are categorized into generic and disease specific (13, 14). Generic PROMs are applicable to any disease and are useful for comparison or economic studies between different conditions. Specific PROMs, on the other hand, are specific for one condition. Both types of measures generic and specific are now seen as complementary rather than conflicting when appraising patient outcomes. Some of the commonly used

generic tools are the European Quality of life 5 dimensions (EQ-5D) (15), the short form 36 (SF-36) questionnaire (16) and the Cleveland Quality of life (17, 18), and the short form 12 (SF-12) (19). Examples of the disease specific PROMs are the Inflammatory bowel disease questionnaire (IBDQ) (20) and the Rating form of IBD patient concerns (21). A good review of PROMs that have been used to evaluate the efficacy of therapeutic agents in functional dyspepsia trials was done by Ang *et al* (22). Mouzas and Pallis reported a good review of the PROMs used in inflammatory bowel disease (23). The patient-reported outcome and quality of life measures database provides a comprehensive list of the available PROM questionnaires (24).

The increasing number of PROMs in the recent years requires gastroenterologists to decide which PROM to use and how to assess each measure. Several studies have suggested using standards to assess properties such as validity, reliability and responsiveness. Examples of these standards have been presented by Terwee *et al* (25), scientific advisory committee of the Medical Outcome Trust (26), EMPRO (Evaluating the Measurement of Patient-Reported Outcomes) tool (27), Bombardier and Tugwell (28), Andresen (29), Steiner (30) Devon *et al* (31), McDowell and Jenkinson (32), the Food and Drug Administration (FDA)(33, 34) and the European Medicines Agency (35) guidelines. The FDA guidance in 2006 (33, 34) describes how to evaluate PROMs used as effectiveness endpoints in clinical trials. These publications describe the required criteria for a successful PROM and are written mainly for health outcome specialists and methodologists who are involved in developing health outcome measures for clinical trials or a evaluating new medical technology. None of these

publications individually summarizes these standards into one relatively brief yet fairly comprehensive practical checklist for doctors to use in their day-to-day clinical practice. Before using health outcome measures for research or in clinical practice, it is essential to ensure they are appropriate to the context, perform well and possess the required characteristics. In this article we describe how to assess these requirements: the concept of items generation, validity, reliability, responsiveness, utility and cross-cultural adaptation; and how to evaluate these measures in a way that is easy to follow and applicable in clinical practice.

THE QUALITY PROPERTIES OF PROMs:

There are 5 main components for good quality PROMs: items selection, validity, reliability, responsiveness and interpretability. With the increase number of multinational and multicultural clinical research, certain criteria regarding cultural, educational and social adaptation of the PROMs are needed to use the questionnaires in a different language or country.

1. Items selection:

Items can be derived from three main sources(13, 36, 37):

1. Research: reviewing old PROMs is the most commonly used approach in finding items. There are several reasons why old measures can be used to derive the new PROMs items; it saves a lot of time and effort, there are possibly a limited number of questions to ask about a specific problem such as abdominal pain, vomiting..etc., and old measures have been repeatedly used and tested in many studies and trials.

2. Patients: by asking the patients to identify items and domains to be included in the scale. Patients can be excellent sources of health outcome measure items. Some techniques like focus groups and key informant interviews have been used to collect patients' viewpoints in a systematic manner (13). This method of item generation has been useful in constructing a quality of life measure for example the inflammatory bowel disease questionnaire (IBDQ)(20), the Rating form of IBD patient concerns (21) and functional dyspepsia (38).
3. Clinical observations: items are derived by clinicians based on their experience.

However, the FDA statement (33, 34) considers including patients in developing PROM questionnaire is the most important source of items. It stresses that items generation should include a wide range of patients to represent variations in severity and in population characteristics such as age, sex, and educational level. It is important to assess the respondents and administrator burden when choosing items. Items that cause undue physical, emotional, or cognitive strain on patients generally decrease the quality and completeness of PROM data. The language in a PROM should be clear and not technical. Items should not offend or discriminate against people for example when assessing the emotional aspect of quality of life. Therefore, items should be tested on small group of patients for a preliminary or a pilot testing to make sure that they are understandable and not ambiguous (39). This pilot testing can include any number of patients. The FDA guidance mention that the number of patients in the pilot testing is not as critical as the cognitive interview quality and patients diversity included in the sample. Pilot

testing of items is commonly used in developing quality of life questionnaires such as the IBDQ and the UK-IBDQ (20, 40).

Once the pool of items has been created, a number of statistical techniques can be used in order to select the most relevant items:

1. Frequency of endorsement: The frequency of endorsement (also called response rate) examines the proportion of people who select the same item response. Only items with endorsement rate between 0.2 – 0.8 (or 20% - 80%) are chosen (13, 41). Items with lower or higher rates are considered redundant because they will add little value to the index. Items with high response rates more than 80% (i.e. more than 80% of patients chose the same answer) are considered for removal because they cannot be used to distinguish between patients. If the same answer was chosen by less than 20% of patients, then it is possibly not related to the condition and can be removed.

2. Item-total correlation: The item-total correlation is the statistical correlation of each item with the total PROM score. The accepted range is between 0.2 to 0.8 (13). A value below 0.2 indicates that the item is not relevant. A value of more than 0.8 indicates that the item is redundant and does not add a value to the total scale.

3. Internal consistency: Internal consistency is the statistical correlation or the homogeneity between the items in the measure (20, 25). The internal consistency is commonly measured by calculating Cronbach α statistic (13, 36) (42). The acceptable value of Cronbach α is between 0.7 and 0.9 (13, 25). Higher values more than 0.9 may indicate an overlap between items.

2.Reliability

Reliability is the consistency between the score of a health outcome measure applied in different circumstances. The principle of reliability is that applying the PROM in different occasions or by different observers produces similar results (13, 43). Statisticians suggest that a reliability of 0.75 should be the minimum requirement for a useful health outcome measure (13). Common reliability statistics are Intraclass Correlation Co-efficient (ICC) and Pearson correlation co-efficient (r). They are expressed as a numbers between -1 and 1 with 0 indicating no reliability; 1 is perfect reliability between the two set of tests and a negative number indicates that the two sets of tests change inversely.

Common types of reliability testing are

- Inter-Observer Reliability is used to assess the degree of consistency between different observers assessing the same patients.
- Test-Retest Reliability is used to assess reliability of the PROM when applied on two separate occasions. To estimate test-retest reliability, the measure should be administered to the same group of patients on two separate occasions between which there has been no overall change in the clinical condition of the patients. Typically a period of 14 days is acceptable (13).

3.Responsiveness:

Responsiveness is the ability of the PROM tool to detect a change in patients' clinical condition. This is estimated by applying the health outcome measure to a group of patients whose clinical condition has changed (13, 44). There are several statistics for responsiveness but the commonest is the

responsiveness ratio, which is calculated by dividing the mean change in scores for patients who reported a change by the standard deviation of the scores of stable patients (13, 44). Other responsiveness indicators mentioned in literature are effect size (ES) (45) and standardized response mean (SRM) (46). The acceptable value for responsiveness ratio is 0.5 or 50% (47) (44, 48).

4. Validity:

Validity is the ability of the test to measure what is intended to measure. Validity can be broadly divided into three types (referred to as the 3 Cs) (13): content validity, construct validity and criterion validity.

Content validity: checks if the measure as a whole covers all the relevant and important aspects of the disease. Experts in the field usually judge content validity to ensure items appropriately measures the desired health outcome.

Construct validity: is used when there is no “gold” standard measure with which a new PROM can be compared (13). A combination of laboratory tests, other health outcome measures, or clinical observations might be necessary to provide the data that support the construct validation of the PROM (25, 49). The common statistic to assess construct validity is by calculating a correlation coefficient. An appropriate correlation coefficient for construct validity should be somewhere between 0.4 and 0.8 (13, 47).

Criterion validity: measures the correlation of the new measure with a “gold standard” measure, which exhibit the same characteristics. When the correlation is explored at the same time then it is described as “*concurrent validation*”. This is often used when an existing measure is potentially to be replaced by a shorter, cheaper or less invasive measure (13). In this case, we

would expect a very high correlation coefficient (≥ 0.8). When the new measure is compared with a criterion that is measured later, this type of validation is called “*predictive validation*”. This type of validation is often used with measures that predict future events like response to treatment or mortality.

5. Interpretability

Interpretability means assigning qualitative meaning to the health outcome measure scores (25, 50-52). To aid using PROM in clinical practice, doctors should be able to translate the PROM score to clinical meaning by knowing the minimal important change (MIC). The minimal important change (MIC) is defined as “the smallest difference in score in the domain of interest which patients perceive as beneficial and would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management” (53). Another useful information is the “Floor and Ceiling Effects”. The ceiling effect is a term used to describe the effect when the majority of the patient scores are close to or at the top of the measure. The floor effect is a term used when the majority of patient scores are close to or at the bottom of the measure (25). A measure is said to have a floor or ceiling effect when more than 15% of patients score the lowest or highest possible score, respectively (25). If floor or ceiling effects are present, it is difficult to accurately assess the health outcomes of patients who score at the extreme ends of the PROM. Results from those groups of patients should be interpreted with caution.

6. Cross-cultural adaptation

Cross-cultural adaptation is the process that deals with language and cultural adaptation issues when preparing the PROMs for use in another country (54, 55). Items should not only be translated linguistically, but also must adapt culturally to the target country culture. For example a questions about difficulty in using fork in eating may not be applicable in a country where fork is not used in eating. The cross cultural adaptation involves forward and backward translation of the questions or items, review the results by linguistics, methodologists, statisticians and health care professionals and pretesting the PROM in the small group of patients to check the clarity of the PROM in the new setting and its consistency with the original PROM version (54). The final step involves psychometric validation of the new PROM in the target population to check validity, reliability, and responsiveness (13, 55). In some ethnic minorities or special group of patients, there is a need for specific cultural and language educational materials such as the use of pictogram, smileys or other picture based representations (56, 57). A good example of a picture based PROM is the Gastro-oesophageal reflux disease analyzer (GERDyzer) questionnaire which comprises 10 dimensions each is illustrated by pictograms drawings (58). In fact a recent study published by Tack *et al* showed that the use of pictograms with verbal descriptors significantly improve the reliability of PWOMs by around 30% by avoiding potential bias by patients (59).

CHECKLIST TO EVALUATE PROMs:

The five important aspects when evaluating the PROMs: Items generation, reliability, validity, responsiveness and interpretability. Each aspect includes one or more psychometric testing. Cross-cultural adaptation is only needed when using the PROM in another language or country. We have produced a simple checklist (Table 1) to evaluate the PROMs without using excessive statistical technical terms.

CONCLUSIONS:

This paper intends to provide a practical overview of the main components for a good quality PROM; it does not intend to provide a detailed description of each component. Readers who require more detailed explanations are encouraged to refer the references cited in the paper. The Food and Drug Administration (FDA)(33) and the European Medicines Agency (35) guidelines provide further recommendations on the proper development and validation of PROMs especially for clinical trials.

A good example of a well-validated PROM is the inflammatory bowel disease questionnaire (IBDQ). IBDQ was developed in 1985 as a quantitative, disease specific HRQoL measure in patients with inflammatory bowel disease. A number of patients with IBD and health professionals were asked to list all problems they had observed or experienced as a direct result of IBD. This process resulted in a total of 150 items. All these items were then administered to another group of patients with IBD to rate each problem on a 5-point Likert scale, ranging from a low score (score 1) indicating no problem to a high score (score 5) indicating a severe problem. A final list of 32 was

derived and reviewed by experienced clinicians, the items were grouped into four groups: Gastrointestinal symptoms, systemic symptoms, emotional dysfunction and social dysfunction (60). The final version of IBDQ had good reproducibility (ICC was 0.7) and responsiveness (by calculating the responsiveness ratio on patients who reported change in a 7-point assessment of their condition). The IBDQ had good construct validity when correlated with disease specific and generic PROMs (61) (62-65). The clinically important change in score was observed to be a decrease of between 16 and 30 points (66). IBDQ was validated into different languages versions and have further proved its validity, internal consistency and reliability in several validation studies worldwide (64, 65, 67-77).

Because new therapies in gastroenterology are rapidly emerging. PROMs are increasingly used in clinical decision-making. There is a need to support and educate gastroenterologist on how to assess these tools to encourage them to use them in clinical practice.

Every PROM tool should have five important properties: Items should come from a reliable source and should be clear to patients, the PROM must yield consistent measurements, i.e. must be reliable, and the PROM instrument must measure the variable it is said to measure, i.e. must be valid, should change with the change in patients' condition "responsive" and can be easily transferred to clinical meaningful values "interpretability".

Conflict of interest: None

Authors' contributions: LA contributed in writing the manuscript, reviewing the literature. HAH and JGW contributed in reviewing the literature and

examining the content of the manuscript. All authors approved the final version of the manuscript

Table 1: Checklist for evaluating the PROM questionnaires:

1. Items Generation: Were items assessed for?
 - Clarity and lack of ambiguity
 - Frequency of endorsement (desired value 20-80%)
 - Internal consistency (desired Cronbach Alpha 0.7-0.9)
 - Item total correlation (desired value 0.2-0.8)
2. Reliability:
 - Was the PROM checked for test-retest reliability
 - Was the PROM checked for inter-observer reliability
 - Was the reliability coefficient acceptable (more than 0.75)
3. Responsiveness:
 - Was the PROM assessed for responsiveness using an appropriate measure (such as responsiveness ratio, effect size, or standardized *response mean* (SRM))
 - Was the value acceptable (e.g responsiveness ratio more than 0.5)
4. Validity:
 - Was the validity of the PROM assessed using the appropriate method (construct validity and/or criterion validity)
 - Was the correlation coefficient \geq the required value of 0.4 for construct validity and 0.8 for criterion validity
5. Interpretability:
 - Can the PROM results be interpreted easily for clinical practice ?
 - Were the values for MIC, SEM, SDC and floor and ceiling effects reported properly?
6. Cross cultural adaptation (only if the PROM is used in a different language/culture):
 - Did the PROM go through a proper cross cultural validation process?

References :

1. Nightingale F. Notes on Hospitals. London: Longman, Green, Longman , Roberts, and Green; 1986.
2. McDowell I. Measuring Health: A Guide to Rating Scales and Questionnaires. 3rd edition ed: Oxford University Press; 2006.
3. World Health Organization. Preamble to the Constitution of the World Health Organization. New York, USA.1948.
4. Khanna P, Agarwal N, Khanna D, Hays RD, Chang L, Bolus R, et al. Development of an Online Library of Patient-Reported Outcome Measures in Gastroenterology: The GI-PRO Database. *American Journal of Gastroenterology*. 2014;109(2):234-48.
5. Devlin NJ, Appleby J. Getting the most out of PROMS: Putting health outcomes at the heart of NHS decision -making. The Kings Fund. London: 2010.
6. Guidance of the Routine Collection of Patient Reported Outcome Measures (PROMs). London: Department of Health. 2007.
7. Lipscomb J, Gotay CC, Snyder CF. Patient-reported outcomes in cancer: a review of recent research and policy initiatives. *CA: a cancer journal for clinicians*. 2007;57(5):278-300.
8. Soreide K, Soreide AH. Using patient-reported outcome measures for improved decision-making in patients with gastrointestinal cancer - the last clinical frontier in surgical oncology? *Frontiers in oncology*. 2013;3:157.
9. Montazeri A. Quality of life data as prognostic indicators of survival in cancer patients: an overview of the literature from 1982 to 2008. *Health and Quality of Life Outcomes*. 2009;7:102.
10. Vandenberghe A, Mion F, Allescher HD, Roman S, Csiki Z. Functional dyspepsia: still a serious challenge for medical practitioners and new drug investigators? A Belgian, French, German and Hungarian opinion. *Acta Gastroenterol Belg*. 2010;73(3):360-5.
11. Irvine EJ, Whitehead WE, Chey WD, Matsueda K, Shaw M, Talley NJ, et al. Design of treatment trials for functional gastrointestinal disorders. *Gastroenterology*. 2006;130(5):1538-51.
12. Drossman DA, Dumitrascu DL. Rome III: New standard for functional gastrointestinal disorders. *Journal of gastrointestinal and liver diseases : JGLD*. 2006;15(3):237-41.
13. Streiner D, Norman G. Health Measurement Scales: A practical guide to their development and use. Fourth ed. Oxford university press 2008.
14. Pallis AG, Mouzas IA. Instruments for quality of life assessment in patients with inflammatory bowel disease. *Dig Liver Dis*. 2000;32(8):682-8.
15. EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. *Health Policy*. 1990;16(3):199-208.
16. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ*. 1992;305(6846):160-4.
17. Fazio VW, O'Riordain MG, Lavery IC, Church JM, Lau P, Strong SA, et al. Long-term functional outcome and quality of life after stapled restorative proctocolectomy. *Annals of surgery*. 1999;230(4):575-84; discussion 84-6.

18. Kiran RP, Delaney CP, Senagore AJ, O'Brien-Ermlich B, Mascha E, Thornton J, et al. Prospective assessment of Cleveland Global Quality of Life (CGQL) as a novel marker of quality of life and disease activity in Crohn's disease. *Am J Gastroenterol.* 2003;98(8):1783-9.
19. Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical care.* 1996;34(3):220-33.
20. Guyatt G, Mitchell A, Irvine EJ, Singer J, Williams N, Goodacre R, et al. A new measure of health status for clinical trials in inflammatory bowel disease. *Gastroenterology.* 1989;96(3):804-10.
21. Drossman DA, Leserman J, Li ZM, Mitchell CM, Zagami EA, Patrick DL. The rating form of IBD patient concerns: a new measure of health status. *Psychosom Med.* 1991;53(6):701-12.
22. Ang D, Talley NJ, Simren M, Janssen P, Boeckxstaens G, Tack J. Review article: endpoints used in functional dyspepsia drug therapy trials. *Aliment Pharmacol Ther.* 2011;33(6):634-49.
23. Mouzas IA, Pallis AG. Assessing quality of life in medical trials on patients with inflammatory bowel disease. *Annals of Gastroenterology.* 2000;13(4):261-3.
24. Patient-Reported Outcome and Quality of life Instruments Database [Internet]. [cited 01/09/2013]. Available from: <http://www.proqolid.org>.
25. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology.* 2007;60(1):34-42.
26. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2002;11(3):193-205.
27. Valderas JM, Ferrer M, Mendivil J, Garin O, Rajmil L, Herdman M, et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2008;11(4):700-8.
28. Bombardier C, Tugwell P. Methodological considerations in functional assessment. *The Journal of rheumatology Supplement.* 1987;14 Suppl 15:6-10.
29. Andresen EM. Criteria for assessing the tools of disability outcomes research. *Archives of physical medicine and rehabilitation.* 2000;81(12 Suppl 2):S15-20.
30. Streiner DL. A checklist for evaluating the usefulness of rating scales. *Canadian journal of psychiatry Revue canadienne de psychiatrie.* 1993;38(2):140-8.
31. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. *Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing / Sigma Theta Tau.* 2007;39(2):155-64.
32. McDowell I, Jenkinson C. Development standards for health measures. *Journal of health services research & policy.* 1996;1(4):238-46.
33. US Food and Drug Administration (FDA). Guidance for industry. Patient reported outcome measures: use in medical product development to support labeling claims. 2006.

34. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health and Quality of Life Outcomes*. 2006;4:79.
35. Szende A, Leidy NK, Revicki D. Health-related quality of life and other patient-reported outcomes in the European centralized drug regulatory process: a review of guidance documents and performed authorizations of medicinal products 1995 to 2003. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2005;8(5):534-48.
36. Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. *Journal of psychosomatic research*. 2010;68(4):319-23.
37. Marx RG, Bombardier C, Hogg-Johnson S, Wright JG. Clinimetric and psychometric strategies for development of a health measurement scale. *Journal of clinical epidemiology*. 1999;52(2):105-11.
38. Carbone F, Holvoet L, Vandenberghe A, Tack J. Functional dyspepsia: outcome of focus groups for the development of a questionnaire for symptom assessment in patients suffering from postprandial distress syndrome (PDS). *Neurogastroenterology and motility : the official journal of the European Gastrointestinal Motility Society*. 2014;26(9):1266-74.
39. Paulsen A, Pedersen AB, Overgaard S, Roos EM. Feasibility of 4 patient-reported outcome measures in a registry setting. *Acta orthopaedica*. 2012;83(4):321-7.
40. Cheung WY, Garratt AM, Russell IT, Williams JG. The UK IBDQ-a British version of the inflammatory bowel disease questionnaire. development and validation. *Journal of clinical epidemiology*. 2000;53(3):297-306.
41. Kovacs FM, Abaira V, Royuela A, Corcoll J, Alegre L, Tomas M, et al. Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain. *BMC musculoskeletal disorders*. 2008;9:43.
42. Kuder GF, Richardson MW. - The theory of the estimation of test reliability. *Psychometrika*. 1937;- 2(- 3):151-60.
43. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of clinical epidemiology*. 2011;64(1):96-106.
44. Sackett DL, Chambers LW, MacPherson AS, Goldsmith CH, McAuley RG. The development and application of indices of health: general methods and a summary of results. *American journal of public health*. 1977;67(5):423-8.
45. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Medical care*. 1989;27(3 Suppl):S178-89.
46. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of clinical epidemiology*. 2008;61(2):102-9.
47. Murphy KaCD. *Psychological Testing Principles and applications*. 4th ed: NJ: Prentice-Hall, Inc; 1998.
48. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *Journal of chronic diseases*. 1985;38(1):27-36.
49. Domino G. *Psychological Testing: An Introduction* 2000, New Jersey: Prentice-Hall Inc.
50. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an

- international Delphi study. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*. 2010;19(4):539-49.
51. Deyo RA, Inui TS. Toward clinical applications of health status measures: sensitivity of scales to clinically important changes. *Health Serv Res*. 1984;19(3):275-89.
 52. Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clinical therapeutics*. 1996;18(5):979-92.
 53. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled clinical trials*. 1989;10(4):407-15.
 54. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000;25(24):3186-91.
 55. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2005;8(2):94-104.
 56. Spaccavento S, Craca A, Del Prete M, Falcone R, Colucci A, Di Palma A, et al. Quality of life measurement and outcome in aphasia. *Neuropsychiatric disease and treatment*. 2014;10:27-37.
 57. Jones CA, Mawani S, King KM, Allu SO, Smith M, Mohan S, et al. Tackling health literacy: adaptation of public hypertension educational materials for an Indo-Asian population in Canada. *BMC public health*. 2011;11:24.
 58. Holtmann G, Chassany O, Devault KR, Schmitt H, Gebauer U, Doerfler H, et al. International validation of a health-related quality of life questionnaire in patients with erosive gastro-oesophageal reflux disease. *Aliment Pharmacol Ther*. 2009;29(6):615-25.
 59. Tack J, Carbone F, Holvoet L, Vanheel H, Vanuytsel T, Vandenberghe A. The use of pictograms improves symptom evaluation by patients with functional dyspepsia. *Aliment Pharmacol Ther*. 2014;40(5):523-30.
 60. Irvine EJ. Development and subsequent refinement of the inflammatory bowel disease questionnaire: a quality-of-life instrument for adult patients with inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 1999;28(4):S23-7.
 61. Russel MG, Pastoor CJ, Brandon S, Rijken J, Engels LG, van der Heijde DM, et al. Validation of the Dutch translation of the Inflammatory Bowel Disease Questionnaire (IBDQ): a health-related quality of life questionnaire in inflammatory bowel disease. *Digestion*. 1997;58(3):282-8.
 62. Kim WH, Cho YS, Yoo HM, Park IS, Park EC, Lim JG. Quality of life in Korean patients with inflammatory bowel diseases: ulcerative colitis, Crohn's disease and intestinal Behcet's disease. *Int J Colorectal Dis*. 1999;14(1):52-7.
 63. Lopez-Vivancos J, Casellas F, Badia X, Vilaseca J, Malagelada JR. Validation of the spanish version of the inflammatory bowel disease questionnaire on ulcerative colitis and Crohn's disease. *Digestion*. 1999;60(3):274-80.
 64. Pallis AG, Vlachonikolis IG, Mouzas IA. Quality of life of Greek patients with inflammatory bowel disease. Validation of the Greek translation of the inflammatory bowel disease questionnaire. *Digestion*. 2001;63(4):240-6.

65. Pallis AG, Mouzas IA, Vlachonikolis IG. The inflammatory bowel disease questionnaire: a review of its national validation studies. *Inflamm Bowel Dis*. 2004;10(3):261-9.
66. Irvine EJ, Feagan B, Rochon J, Archambault A, Fedorak RN, Groll A, et al. Quality of life: A valid and reliable measure of therapeutic efficacy in the treatment of inflammatory bowel disease. *Gastroenterology*. 1994;106(2):287-96.
67. Muller S, Jan Irvine E, Gathany T. PGI18 LINGUISTIC VALIDATION OF THE INFLAMMATORY BOWEL DISEASE QUESTIONNAIRE (IBDQ) IN 35 LANGUAGES. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2008;11(3):A89.
68. Verissimo R. Quality of life in inflammatory bowel disease: psychometric evaluation of an IBDQ cross-culturally adapted version. *Journal of gastrointestinal and liver diseases : JGLD*. 2008;17(4):439-44.
69. Hjortswang H, Jarnerot G, Curman B, Sandberg-Gertzen H, Tysk C, Blomberg B, et al. Validation of the inflammatory bowel disease questionnaire in Swedish patients with ulcerative colitis. *Scand J Gastroenterol*. 2001;36(1):77-85.
70. Bernklev T, Moum B, Moum T. Quality of life in patients with inflammatory bowel disease: translation, data quality, scaling assumptions, validity, reliability and sensitivity to change of the Norwegian version of IBDQ. *Scand J Gastroenterol*. 2002;37(10):1164-74.
71. Hashimoto H, Green J, Iwao Y, Sakurai T, Hibi T, Fukuhara S. Reliability, validity, and responsiveness of the Japanese version of the Inflammatory Bowel Disease Questionnaire. *J Gastroenterol*. 2003;38(12):1138-43.
72. Pontes RM, Miszputen SJ, Ferreira-Filho OF, Miranda C, Ferraz MB. [Quality of life in patients with inflammatory bowel diseases: translation to Portuguese language and validation of the "Inflammatory Bowel Disease Questionnaire" (IBDQ)]. *Arq Gastroenterol*. 2004;41(2):137-43.
73. Janke KH, Klump B, Steder-Neukamm U, Hoffmann J, Hauser W. [Validation of the German version of the Inflammatory Bowel Disease Questionnaire (Competence Network IBD, IBDQ-D)]. *Psychother Psychosom Med Psychol*. 2006;56(7):291-8.
74. Masachs M, Casellas F, Malagelada JR. [Spanish translation, adaptation, and validation of the 32-item questionnaire on quality of life for inflammatory bowel disease (IBDQ-32)]. *Rev Esp Enferm Dig*. 2007;99(9):511-9.
75. Ren WH, Lai M, Chen Y, Irvine EJ, Zhou YX. Validation of the mainland Chinese version of the Inflammatory Bowel Disease Questionnaire (IBDQ) for ulcerative colitis and Crohn's disease. *Inflamm Bowel Dis*. 2007;13(7):903-10.
76. Vidal A, Gomez-Gil E, Sans M, Portella MJ, Salamero M, Pique JM, et al. Psychometric properties of the original Inflammatory Bowel Disease Questionnaire, a Spanish version. *Gastroenterol Hepatol*. 2007;30(4):212-8.
77. Ciccocioppo R, Klersy C, Russo ML, Valli M, Boccaccio V, Imbesi V, et al. Validation of the Italian translation of the Inflammatory Bowel Disease Questionnaire. *Dig Liver Dis*. 2011;43(7):535-41.