Development of the BD-QoL: A Quality of Life Measure Specific to Behçet's Disease

GILL GILWORTH, M. ANNE CHAMBERLAIN, BIPIN BHAKTA, DORIAN HASKARD, ALAN SILMAN, and ALAN TENNANT

ABSTRACT. *Objective.* Current outcome measures for patients with Behçet's disease (BD) are impairment-focused and do not necessarily take account of the wider impact of the condition on the individual's lifestyle. Our aim was to develop a disease-specific measure of quality of life (QoL) for BD.
Methods. The content of the BD-QoL was derived from qualitative interviews with patients using a "needs-based" approach to identify items. A postal survey was used to test the scaling properties, reliability, internal consistency, and validity of the new questionnaire using Rasch analysis. A second postal survey was used to assess test-retest reliability and internal consistency and to provide further evidence of the validity of the questionnaire.
Results. Main themes emerging from the qualitative interviews included relationships, emotions, limitations in day to day activities, and self-image. From these themes 71 statements were chosen as potential items for the BD-QoL. After analysis, 30 items of the BD-QoL emerged free of item bias for age and sex. Fit to the Rasch model was excellent. In the second postal survey test-retest reliability of the 30 item BD-QoL was 0.84.
Conclusion. The BD-QoL provides the clinician with a simple, reliable, and valid tool for assessing the influence of interventions for BD and for evaluating models of service delivery. It is well accepted by patients, and has excellent scaling and psychometric properties. The BD-QoL complements information obtained through BD-specific disease activity scales. (J Rheumatol 2004;31:931–7)

Key Indexing Terms:
BEHÇET'S DISEASE   OUTCOME   QUALITY OF LIFE   QUESTIONNAIRE

Behçet's disease (BD) is a chronic multisystem disorder characterized by vasculitis affecting small to medium size blood vessels. Although the commonest symptoms of BD relate to mucosal ulceration it can affect virtually every organ system, causing a variety of clinical problems such as arthritis, neurological impairments, pulmonary artery aneurysm, and gastrointestinal symptoms. The commonest severe manifestation is uveitis, which occurs in as many as 70% of patients. There is no uniformly accepted laboratory test for diagnosis, and therefore classifying people as having BD depends on the presence of internationally agreed criteria[1] based on clinical features.

The complex pattern of signs and symptoms in BD can lead to a variety of activity limitations (disability) and restriction in participation in many areas of life (handicap). A dilemma thus exists of the choice of outcome measure(s) most likely to identify change as a consequence of clinical intervention. To date the measurement of outcome in BD has mainly focused on impairments. The absence of laboratory markers that correlated well with impairment led to the development of a standardized proforma (the BD Current Activity Form) to assess disease activity that is based on history of specific clinical features[2]. This is an impairment-based outcome measure useful in monitoring therapy. However, it takes no account of the wider impact of the condition on the individual's lifestyle.

Generic outcome measures of health status such as the Nottingham Health Profile (NHP)[3] and the Medical Outcomes Study Short-Form 36 (SF-36)[4] can appear attractive as tools for measuring the influence of disease because they often measure several domains and allow for comparisons between different conditions. However, condition-specific measures have been shown to be more sensitive to change[5]. Health utility measures, such as the EUROQoL, tend to be even more restricted, as they include only a few items concerning impairments and activity limitation and

this can lead to inconsistency in the health state evaluations[6]. In an attempt to overcome the limitations of generic outcomes measures, some have been modified to account for disease-specific attributes. An example is the use of a modified version of the Stanford Health Assessment Questionnaire (HAQ) in patients with psoriatic arthritis to take account of skin disease[7]. Another is the adaptation of the SF-36 into the MSQoL54 for multiple sclerosis, which adds new items to existing domains, as well as new domains[8].

There has been growing interest in the assessment of quality of life (QoL), particularly in chronic disabling conditions. The importance of the construct was recognized by the UK National Health Service *Review Working Paper on Medical Audit* that saw QoL as an integral part of the audit procedure[9]. QoL assessment is also becoming routine in clinical trials. While outcome measures that focus on symptoms and function provide important information about the level of impairment or activity limitation experienced by the patient, they do not necessarily inform on the influence of the condition on the patient's life. The measurement of QoL can be conceived as a summary of this influence, together with the effect of the treatment, on the patient[10]. As such, it is a more useful construct for evaluating the outcome of interventions, particularly with conditions presenting with such diverse and complex patterns of impairment. This is because this approach goes beyond the impairment–activity-participation continuum[11] by asking what the patients' health prevents them from doing and also about their emotional response to these restrictions[12]. Consequently, QoL represents a substantively different construct from the impairments and activity-limitations domains assessed by the outcome measures currently used with BD.

One widely used conceptualization of QoL is the "needs-based" model proposed by Hunt and McKenna[13], which draws on theories of human motivation[14]. The model postulates that individuals are motivated or driven by their needs that are either inborn or learned during socialization processes. The model is based on the premise that life gains its quality from the ability and capacity of the individual to satisfy certain human needs. QoL is high when most human needs are fulfilled and low when few needs are satisfied. The model has been used in the development of several condition-specific QoL instruments. These include measures for depression[15], adult growth hormone deficiency[16], recurrent genital herpes[17], migraine[18], rheumatoid arthritis (RAQoL)[19], and ankylosing spondylitis[20]. The content of these instruments was derived from a validated technique involving in-depth qualitative interviews with relevant patients. In this way the needs relevant to each condition were identified, maximizing the content validity and responsiveness of the final instruments. For example, the RAQoL has high test-retest reliability (> 0.90) and construct validity, and is now widely used[19,21].

One criticism of disease-specific QoL measures has been that they do not allow comparisons across diseases for economic analyses, or provide a common measurement system (for example, like the HAQ or health utility scales) with which clinicians and others can become familiar. However, recent advances in the evaluation of outcome measures using techniques such as Rasch analysis (see below) have presented the possibility of overcoming this criticism by providing the ability to combine individual disease-specific QoL measures on a single common metric scale. This allows comparisons between different disease groups while maintaining the relevance and responsiveness of the instruments used to summarize the effect of the disease on an individual.

MATERIALS AND METHODS

Patients. Patients who fulfilled the International Study Group Classification Criteria for BD were recruited from 3 hospitals, 2 in the North of England and one in London. Different patients were involved at each stage of the study and patients with significant comorbidity such as primary mental health problems were excluded.

Development Methodology
The stages involved in the development of the BD-QoL are set out in Figure 1.

Stage 1: Qualitative interviews with patients. The content of the BD-QoL was derived from qualitative interviews with patients, conducted by 2 experienced qualitative interviewers. Thus specific statements included within the scale are derived from the patients' own comments, and often in their own words. This approach ensures that the content of the final instrument is relevant to the target population and that issues considered important by patients are not omitted. A theoretical sample frame was constructed with patients stratified according to age (< 40 years of age and > 40 years of age), sex, and main presenting symptoms (Table 1).

The interviews took the form of informal, focused conversations designed to explore the influence of BD on the life of the patient. The interviewers carried out the interviews at the subject's home, place of work, or at the outpatient clinic, depending on individual preferences. All subjects were given a choice of sex of interviewer. With the permission of the patient the interviews were tape-recorded. Typed transcripts were produced from these tapes. To maintain anonymity, any information that would allow identification of the patient was omitted from the transcripts. A sample of the transcripts was read by 4 members of the research team plus an independent researcher in order to identify agreement on the main themes emerging. Detailed analysis of all the interview transcripts was carried out using grounded theory principles[22] with the assistance of NUD*IST computer software[23].

Stage 2: Selection of items for the draft questionnaire. The transcripts were analyzed to identify potential items for the questionnaire. Items selected fulfilled the following criteria: (1) relevant to the needs model of QoL, (2) applicable to all potential respondents, and (3) reflected a single idea and were unambiguous.

As far as possible, potential items selected for the questionnaire were direct quotations from the transcripts. However, the actual words used by patients were altered for some items so that they were expressed in the first person and/or in the present tense. A draft questionnaire was prepared using the items identified. The response format was a dichotomous "true/false," as this has been found most appropriate and robust for these types of statement[19].

Stage 3: Field-testing for face and content validity (cognitive debriefing). This stage was designed to test whether patients with BD were able to understand and complete the draft questionnaire, to ensure that the content
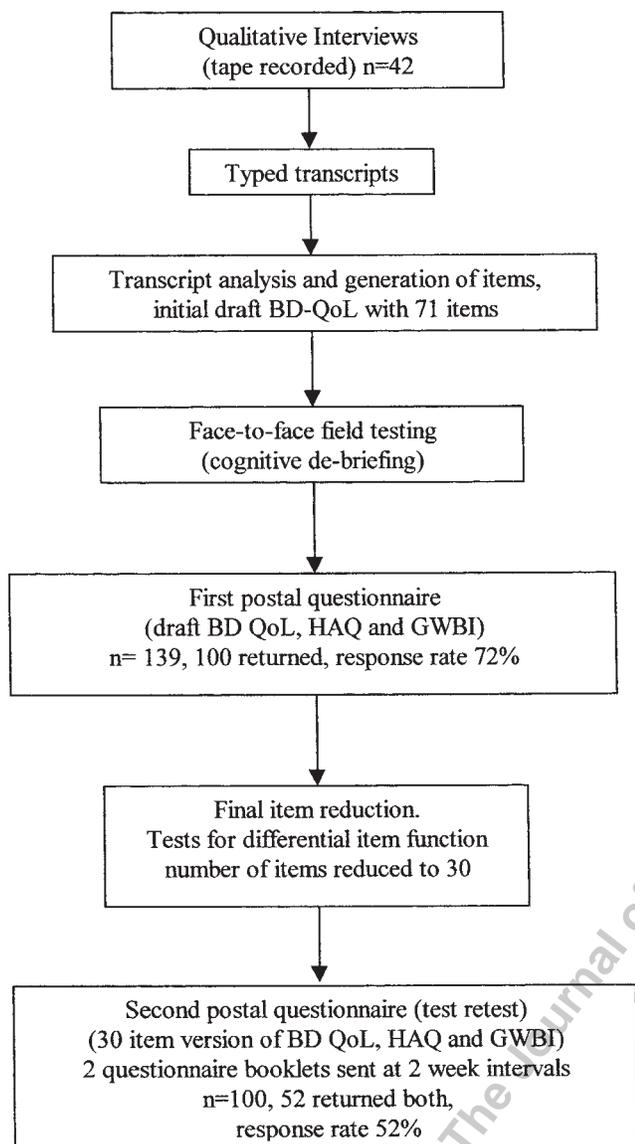
Figure 1. Stages of the study.

was relevant and that important issues had not been omitted. Patients were approached in a dedicated BD outpatient clinic. Those who agreed to participate were interviewed either before or after their consultation. Interviewees were asked to complete the draft BD-QoL in the presence of an interviewer and then to comment on its ease of completion and on the appropriateness of the instructions, items, and response format. Items found

to be problematic were reworded. As a result of this exercise, a second draft version of the measure was prepared.

Stage 4: Initial postal survey. The purpose of the first postal survey was to test the scaling properties of the draft BD-QoL, to facilitate item reduction, and to provide preliminary evidence of construct validity. Patients were contacted and asked to complete and return a questionnaire booklet. This consisted of demographic questions, the draft BD-QoL, the HAQ, and the General Well Being Index[24]. The latter is a generic measure of subjective well-being.

Rasch analysis was conducted on the data to identify items that failed to fit the underlying measurement construct (QoL) and/or that worked differentially by age or sex. The application of the Rasch model ensures that the fundamental properties of the instrument (for example, unidimensionality and level of measurement) are assessed, in addition to the traditional psychometric assessments of reliability and external construct validity[25]. In the present context, the Rasch model adopts the premise that the likelihood of a person affirming a particular item depends on both the level of QoL of the person and on the level of QoL represented by the item. The model assumes that the probability of a given respondent affirming an item is a logistic function of the relative distance between the item location and the respondent location on a linear scale. In other words the probability that a person will affirm an item is a logistic function of the difference between the person's QoL [θ] and the level of QoL expressed by the item [b], and only a function of that difference.

$$p_{ni} = \frac{e^{(\theta n - b_i)}}{1 + e^{(\theta n - b_i)}}$$

where p_{ni} is the probability that person n will affirm the item i, θ is person QoL, and b is the level of QoL expressed by the item. The formula can be expressed as a logit model:

$$\ln \left(\frac{p_{ni}}{1 - p_{ni}} \right) = \theta_n - b_i$$

where P is the probability of person n affirming item i, θ is person QoL, and b is the level of QoL expressed by the item. Fitting data to the Rasch model thus places both item and person parameter estimates on the same log-odds units (logits) scale.

Statistics indicating fit to the model test how far the observed data match that expected by the model. Three overall fit statistics are considered. Two are item-person interaction statistics distributed as a Z statistic with mean of zero and standard deviation of 1 (which indicates perfect fit to the model). A third is an item-trait interaction statistic reported as a chi-square, reflecting the property of invariance across the trait. A significant chi-square indicates that the hierarchical ordering of the items varies across the trait. In addition, individual person- and item-fit statistics are presented, both as residuals (a summation of individual person and item and deviations) and as a chi-square statistic (deviation from the model by groups of people). In this study, a value of 0.01 was defined as the significance level for individual item-fit, to take account of multiple testing (of items). Residuals between ± 2.5 were deemed to indicate adequate fit to the model.

Table 1. Sampling frame for Behçet's disease quality of life qualitative interviews. Numbers indicate the subjects interviewed. The aim was to have 3 or 4 interviewees in each cell.

Main Presentation of Disease	< 40 Years of Age		> 40 Years of Age	
	Male	Female	Male	Female
Ulceration/arthritis/skin involvement	6	5	3	4
Eye involvement	3	6	2	1
Other major systems, e.g., stroke or other neurological symptoms	1	2	5	4

In addition, differential item functioning (DIF) for age and sex is examined in detail[26]. Under the requirement that the ability under consideration is unidimensional, if an item measures the same level of QoL across groups, then, except for random variations, the item should work in the same way irrespective of the nature of the group (e.g., sex) for whom a function is plotted[27]. Here the number of tests for DIF undertaken was large (e.g., every item by age and sex) and thus the significance level was set at 0.001 to accommodate repeated tests.

Stage 5: Postal survey 2. The purpose of the second postal survey was to determine the scaling properties, internal consistency, reliability, and construct validity of a revised version of the BD-QoL, following the analysis from Stage 4. Patients were sent a new questionnaire booklet. Patients who completed and returned this were sent a second booklet 2 weeks later.

Once again, data from the BD-QoL were fitted to the Rasch model to confirm that the items formed a unidimensional scale, that they each represented a different amount of QoL (hierarchical ordering), and to confirm the absence of DIF. The analyses also determined whether the scale operated at the ordinal or interval level. Internal consistency (the degree of relatedness between items) was assessed by Cronbach's alpha coefficient. A value of 0.70 or above was taken as being indicative of adequate internal consistency.

The reliability of the BD-QoL (an estimate of the instrument's reproducibility over time, assuming that no change in condition has taken place) was assessed using the test-retest method. Scores on the instrument at each administration were correlated.

The study was granted local research ethics committee approval by the United Leeds Teaching Hospitals Trust, Manchester Royal Infirmary, and Hammersmith Hospitals NHS Trust.

Sample size and statistical software. Rasch analysis requires a minimum sample size of 45 cases per group to test for DIF (e.g., by sex) where, at 95% significance, a difference of 0.6 standard deviations within the residuals can be detected with 80% power[28]. For overall parameter estimation with a reasonably targeted sample of 50 persons there is a 99% confidence that the estimated item difficulty is within ± 1 logit of its stable value, which is close enough for most practical purposes, especially when participants respond to 10 or more items[29].

Qualitative text was analyzed using the NUD*IST package, version 4. Quantitative data analysis was undertaken with SPSS version 9; Rasch analysis with RUMM2010[30].

RESULTS

Findings from qualitative interviews. In total, 42 subjects fulfilling the criteria and matching the theoretical sampling frame agreed to be interviewed out of a total of 64 approached (66%). Although the subjects were not evenly distributed throughout the sampling frame, at least one subject was interviewed in each of the cells (Table 1). The age range of the sample was 20–66 years (mean 41.3), and 52% were female.

For the qualitative interviews, 4 female subjects and one male subject requested a female interviewer; none specified a preference for the male interviewer. Some of the issues raised by the interviewees were common to many chronic rheumatological conditions, including limitations in daily activities, the effect on relationships, and self-confidence. However, other issues appeared to characterize the experience of BD more specifically; for example, problems with talking and eating for patients with mouth ulceration. A whole range of needs are frustrated as a result of having BD and almost all the subjects spoke of the effect of their illness on their moods and emotions. In addition, many subjects reported their illness has influenced large parts of their life including career choices, family planning, and even where they chose to live.

Selection of items for the draft questionnaire and field-testing for face and content validity (cognitive debriefing). An item selection process that identified statements consistent with an expression of need led to 71 potential items. This 71 item draft BD-QoL was then field-tested with 6 BD patients in an outpatient clinic. The questionnaire was well accepted by interviewees, who generally found the items to be understandable and relevant. No significant problems were found in their ability to complete the questionnaire, although simplified wording of 3 of the items was suggested and adopted at this stage.

Results of the first postal survey testing the psychometric and scaling properties of the BD-QoL. Questionnaire booklets were then distributed to 139 subjects and returned by 100, a response rate of 72%. The mean age of patients who returned the booklet was 43 years (SD 11.6); 59% were female, and overall, respondents reported a mean duration of disease of 16.6 (SD 19.4) years. More than 4 in 5 (81%) reported that they lacked the energy to do usual activities, and 45% reported that they had experienced these symptoms for 2 or more weeks. Fifty-nine percent reported headaches (33% for 2+ weeks); 57% mouth ulcers (22%); 25% genital ulcers (7%); 52% skin problems (28%); 73% joint pain or stiffness (48%); and 57% reported stomach problems (33%). Almost two-fifths (37%) reported eye symptoms. The mean HAQ was 0.64 (SD 0.069). Just less than one-third of patients (31%) reported that, compared to how they usually were, today was a "good day." Well over half (57%) reported that it was an "average day" and the remainder (12%) that it was a "bad day." There was a strong association between this variable and a variable recording the overall perception of disease, a 7 "smiley face" question taken from the Disease Activity Form2. In this question, patients are requested to indicate how they have been feeling over the last 4 weeks thinking only about their BD. The response options are a row of 7 faces with a range of facial expressions ranging from very happy to very sad. Patients reporting the 3 most smiling faces never reported a "bad day" (chi-square 31.9, p = 0.001). In total, almost half (46%) reported the 3 "sad" faces for disease activity, and mean overall health (by visual analog scale, VAS) was 49.4 (SD 24.7).

After Rasch analyses, 30 of the BD-QoL items emerged as belonging to a unidimensional QoL scale and free from differential item functioning for age and sex. Fit to the Rasch model of these 30 items was excellent. Mean item fit was 0.192 (SD 0.58) and person fit 0.253 (SD 0.84). Item-trait interaction chi-square significance was 0.64, showing invariance across the QoL trait. Individual item fit chi-square ranged from 0.97 to 0.02. Individual residual fit ranged from –1.26 to 1.29. The best and worst fitting of these 30 items are shown in Table 2. The location of the item represents its frequency and impact

Table 2. Fit to the Rasch model. Best and worst fitting items (chi-square) from the 30 item BD-QoL. (Full copies of the scale are available upon request from the author.)

Item	Location	Residual	Chi-square Probability
Best fitting			
It is difficult to get out of bed	–0.175	1.293	0.971
My life revolves around hospital visits	0.367	0.371	0.922
My condition is drastically affecting my life	–0.172	–0.720	0.900
I often get frustrated	–2.286	–0.070	0.889
My condition affects important decisions in my life	–1.341	–0.187	0.807
Worst fitting			
I have lost contact with people	0.379	–0.826	0.231
I feel lonely	0.594	0.794	0.117
I feel dependent upon others	0.260	–0.012	0.092
I feel unable to cope with my condition	1.990	–1.123	0.072
I feel older than my years	–1.224	–0.018	0.023

upon QoL. The Rasch analysis arbitrarily assigns the item of average impact a location of zero and so the scale is centered on the average item. Those items with a high negative location are common, and mark the lower part of the QoL construct (better QoL). Those with high positive locations are rare, and are associated with high scores (worst QoL). For example, "I often feel frustrated" has a high negative location and is thus common, whereas affirmation of the item "I feel unable to cope with my condition," which has a high positive location, is rare. People who respond positively to this item are likely to have a poor QoL (high score).

How patients responded to the "Compared to how you usually are, is today?" question was found to be closely associated with their QoL. Mean BD-QoL score for those reporting a good day was 8.6 (SD 6.5), an average day 12.4 (8.0), and a bad day 21.5 (7.5) (F = 12.51, p < 0.001). The correlation between the "smiley faces" disease perception and the BD-QoL was, as expected given the conceptual model, 0.6. Thus disease activity, largely predicated upon impairment with some activity limitation, explains about 36% of the variation in QoL.

Results of the second postal survey. The 30 item draft BD-QoL was sent to a further sample of 100 subjects. Fifty-two completed and returned the questionnaire on 2 occasions, 2 weeks apart. Mean age of responders was 44.2 years (SD 10.6) and 77% were female. For this group, mean duration of disease was 12.6 (SD 10.6) years and mean health (by VAS) was 52.2 (SD 27.0). On completion of their first questionnaire 27% reported that it was a good day, 53% an average day, and 20% a bad day.

Once again, fit of the 30 items to the Rasch model was found to be satisfactory. For the second response in the test-retest study, item fit was –0.22 (SD 0.87) and person fit –0.22 (SD 0.98) with chi-square interaction fit statistic showing 36.43 (df 30) with significance of p = 0.19. Items were found to be unequally spaced along the measurement continuum, indicating that the 30 item BD-QoL produces raw scores at the ordinal level of measurement. Finally, person-separation reliability was shown to be 0.943, indicating that the scale is able to differentiate 4 or more groups of people among the continuum[31].

Test/retest reliability of the 30 item BD-QoL was 0.84. Internal consistency (alpha) of the 30 item scale was 0.94. Adequate external construct validity was demonstrated by a correlation of 0.6 with the HAQ and –0.76 with the General Well Being Index, thus confirming the expected moderate association with activity limitation and strong association with subjective well-being. The Kolmogorov-Smirnov Z test significance of 0.315 confirmed a normal distribution of the new scale.

Once again a significant association was found between the BD-QoL and the "today" question (F = 9.01, p < 0.001), and the association between QoL and disease activity (smiley faces) was found to be 0.7 on this occasion.

DISCUSSION

Instruments currently available for use with patients with BD tend to focus on symptoms (impairment) and are used to assess the presence or absence of disease and its consequences in these terms. Although such measures provide important information on the level of impairment experienced by patients, they do not necessarily inform on the influence of the condition on QoL. QoL goes beyond the impairment, activity, and participation continuum by asking what the patients' health prevents them from doing and also about their emotional response to these restrictions. The development of the BD-QoL and resulting analysis confirms that disease activity had only a moderate association with QoL where the latter is defined through a needs-based approach. This is important, as application of the BDQoL alongside the BD current activity form (in the context of routine clinical practice and within research trials) provides additional information about the patient. The BDQoL is easy to complete and does not require additional time within the clinical consultation to

complete given that it is a self-completed questionnaire. The use of the sampling frame (Table 2) ensures that the items within the BDQoL scale correctly reflect the QoL issues that may arise in the spectrum of patients that one would see in clinical practice.

Increasingly, Rasch analysis is seen as the "gold standard" for constructing measures, with its emphasis on delivering fundamental measurement to the health sciences[31,32]. Laying down the requirements of measurement to support arithmetic operations such as the calculation of change scores, it adds a scientific basis to internal construct validity and the scaling properties of instruments[32,33]. This supplements the traditional psychometric analysis concerning reliability and validity, and thus offers an invaluable tool for constructing measures. Further, through its calibration onto an underlying metric, the potential for co-calibration with other QoL scales (by common-item equating) offers the scope for comparable disease-specific QoL measures through "item banking"[34].

Taken together with the theoretical construct of QoL put forward by the needs-based model, a powerful scientific approach to measurement can thus be utilized to provide fundamental measurement for health outcomes. We have demonstrated that the BDQoL derived from these approaches delivers quality measurement in terms of classical as well as modern psychometric theory. However, as with all developments of new scales, further construct validation will be required, as well as studies of responsiveness.

We believe the BD-QoL will be a valuable tool for assessing the influence of BD and its treatment on QoL in clinical settings and research studies. Such an instrument will allow accurate assessment of the effectiveness of interventions from the patient's perspective. The psychometric and scaling properties of the BD-QoL suggest that researchers and clinicians can have confidence in the scores obtained by respondents on the measure. Further assessments of the instrument's validity and responsiveness will be required. Adaptations into other languages will follow, and will conform to current adaptation guidelines[35].

ACKNOWLEDGMENT

REFERENCES

1. International Study Group for Behçet's Disease. Evaluation of diagnostic (classification) criteria in Behçet's disease — Towards internationally agreed criteria. Br J Rheumatol 1992;31:299-308.
2. Bhakta BB, Brennan P, James TE, Chamberlain MA, Noble BA, Silman AJ. Behcet's disease: a new instrument to measure clinical activity. Rheumatology Oxford 1999;38:728-33.
3. Hunt SM, McEwan J, McKenna SP. Measuring health status: a new tool for clinicians and epidemiologists. J Roy Coll Gen Pract 1985;35:185-8.
4. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Med Care 1992;30:473-83.
5. Dawson J, Carr A. Outcome evaluation in orthopaedics. J Bone Joint Surg Br 2001;83:313-5.
6. Doland P, Kind P. Inconsistency and health state valuations. Soc Sci Med 1996;42:609-15.
7. Husted JA, Gladman DD, Long JA, Farewell VT. A modified version of the Health Assessment Questionnaire for psoriatic arthritis. Clin Exp Rheumatol 1995;13:439-43.
8. Vickrey BG, Hays RD, Harooni R, Myers LW, Ellison GW. A health-related quality of life measure for multiple sclerosis. Qual Life Res 1995;4:187-206.
9. National Health Service Review. Working paper No. 6, medical audit. London: HMSO; 1989.
10. Tennant A, McKenna SP. Conceptualising and defining outcome. Br J Rheumatol 1995;34:899-900.
11. WHO. International classification of functioning, disability and health. Geneva: World Health Organisation; 2001.
12. McKenna SP, Hunt SM, Tennant A. Psychological well-being in depressed patients. Int J Methods Psych Res 1993;3:245-51.
13. Hunt SM, McKenna SP. The QLDS: A scale for the measurement of quality of life in depression. Health Policy 1992;22:307-19.
14. Maslow AH. Motivation and personality. 2nd ed. New York: Harper & Row; 1970.
15. McKenna SP, Hunt SM. A new measure of quality of life in depression: Testing the reliability and construct validity of the QLDS. Health Policy 1992;22:321-30.
16. Holmes SJ, McKenna SP, Doward LC, Shalet SM. Development of a questionnaire to assess the quality of life in adults with growth hormone deficiency. Endocrinol Metab 1995;2:63-9.
17. McKenna SP, Doward LC. Measuring quality of life in recurrent genital herpes. Eur Rev Appl Psychol 1995;45:289-95.
18. McKenna SP, Patrick DL. The development of a migraine-specific quality of life measure for multi-national clinical trials [abstract]. Cephalalgia 1995;15 Suppl 14:280.
19. De Jong Z, van der Heijde D, McKenna SP, Whalley D. The reliability and construct validity of the RA-QoL. Br J Rheumatol 1997;36:878-83.
20. Doward LC, Spoorenberg A, Cook SA, et al. The development of the ASQoL: A quality of life instrument specific to ankylosing spondylitis. Ann Rheum Dis 2003;62:20-6.
21. Kutlay S, Kucukdeveci AA, Gonul D, Tennant A. An adaptation and validation of the Turkish version of the Rheumatoid Arthritis Quality of Life Scale. Rheumatol Int 2003;23:21-6.
22. Strauss A, Corbin J. Basics of qualitative research — Grounded theory procedures and techniques. London: Sage Publications; 1990.
23. Fielding NG, Lee RM, editors. Using computers in qualitative research. London: Sage Publications; 1991.
24. Hunt SM, McKenna SP. British adaptation of the General Well-Being Index: a new tool for clinical research. Br J Med Econ 1992;2:49-60.
25. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1960 (reprinted 1980).
26. Angoff WH. Perspectives on differential item functioning methodology. In: Holland PW, Wainer H, editors. Differential item functioning. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993.
27. Dorans NJ, Holland PW. DIF detection and description: Mantel-Haenszel and standardisation. In: Holland PW, Wainer H, editors. Differential item functioning. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993:36-66.
28. Elashoff JD. NQuery Advisor. Version 4.0. User's guide. Boston: Statistical Solutions; 2000.
29. Wright BD, Tennant A. Sample size again. Rasch Measurement

Transactions 1996;9:468.

30. Andrich D, Lyne A, Sheridon B, Luo G. RUMM 2010. Perth: RUMM Laboratory; 2000.

31. Fisher WP. Reliability statistics. Rasch Measurement Transactions 1992;6:238.

32. Andrich D. Rasch models for measurement. London: Sage Publications; 1988.

33. Svensson E. Guidelines to statistical evaluation of data from rating scales and questionnaires. J Rehab Med 2001;33:47-8.

34. Dobby J, Duckworth D. Objective assessment by means of item banking. Schools Council Examination Bulletin 40. London: Evans/Methuen Educational; 1979.

35. Beaton D, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 2000;25:3186-91.