Expert evaluations of fatigue questionnaires used in rheumatoid arthritis: a Delphi study among patients, nurses and rheumatologists in the Netherlands

S. Nikolaus[1], C. Bode[1], E. Taal[1], M.A.F.J. van der Laar[1,2]

[1]IBR Research Institute for Social Sciences and Technology, University of Twente; [2]Department of Rheumatology and Clinical Immunology, Medical Spectrum Twente, Enschede, The Netherlands.

Abstract

Objectives

Evaluating fatigue items from traditional questionnaires and a new scale (BRAF-MDQ) by experts in rheumatoid arthritis (RA). This evaluation was part of a study to select fatigue items to develop an item bank for a Dutch computer-adaptive test (CAT) for RA. Experts' opinions were incorporated since they are essential for content validity of measurement instruments.

Methods

The 60 items of the SF-36 subscale vitality, FACIT-F, POMS subscale fatigue/inertia, MAF and the recently developed BRAF-MDQ were evaluated by rheumatologists, nurses and RA patients in a Delphi procedure. Items were selected for development of the item bank/CAT if rated as adequate by at least 80% of the participants (when 50% or less they were excluded). On the basis of participants' comments, remaining items were re-worded and re-evaluated in the following round. The procedure stopped when all items were selected or rejected.

Results

Ten rheumatologists, 20 nurses and 15 RA patients participated. After the first round, 40% of the traditional items and 60% of the BRAF-MDQ items were directly selected and 3 items of the traditional questionnaires and 1 item of the BRAF-MDQ were directly excluded. Remaining items were re-worded, eight of which were presented for re-evaluation in the second round. Finally, 90% of the items from the traditional questionnaires and 95% of the items from the new BRAF-MDQ were included in our item pool.

Conclusion

Fifty-five of the 60 items (92%) from fatigue questionnaires proved to have good content validity and were feasible for use in the Netherlands, some after adaptation.

Key words

fatigue, rheumatoid arthritis, Delphi technique, expert opinion, patient outcome assessment

Stephanie Nikolaus, MSc
Christina Bode, PhD
Erik Taal, PhD
Mart A.F.J. van der Laar MD, PhD

Please address correspondence to:
Stephanie Nikolaus,
University of Twente
Drienerlolaan 5
7500 AE Enschede, The Netherlands
E-mail: s.nikolaus@utwente.nl

Received on April 29, 2011; accepted in revised form on September 20, 2011.

Funding: This study was financed by the Foundation of Rheumatism Research Twente (Stichting Reumaonderzoek Twente) and the IBR Research Institute for Social Sciences and Technology, and conducted at the Arthritis Centre Twente; a collaboration between the University of Twente and Medical Spectrum Twente.

Competing interests: none declared.

Introduction

Measuring fatigue in rheumatoid arthritis (RA) provides information on a prominent symptom of RA, and is essential for the understanding of the patients' perspective (1). Patients experience fatigue as a multidimensional, annoying symptom with far-reaching consequences (2-5).

Only four multi-item fatigue questionnaires have shown reasonable evidence for validity in RA (6): Short Form 36 subscale vitality (SF-36; 7), Functional Assessment of Chronic Illness Therapy Fatigue Scale (FACIT-F; 8), Profile of Mood States subscale fatigue/inertia (POMS;9), and the Multidimensional Assessment of Fatigue scale (MAF; 10). However, none of these scales met all criteria for validity nor was the perspective of patients included in their development (6). The Bristol RA Fatigue Multi-Dimensional Questionnaire (BRAF-MDQ; 11) has been recently constructed and evaluated in a British RA population. Its development encompasses the patients' perspective with interviews, focus groups and cognitive testing. The development focuses on patients scoring 7 or higher on a VAS fatigue and the evaluation of the questionnaire on patients scoring 5 or higher (2, 11).

There is still no measurement instrument that incorporates the patients' perspective and that is validated in the Netherlands. In this study, the BRAF-MDQ items will be tested in a Dutch population for the first time. However, we do not intend to develop another fatigue questionnaire; instead we will use innovative technology to measure fatigue in RA more precisely with fewer items. Computer-adaptive testing (CAT) allows comprehensively measuring fatigue with relatively few items (12). Items are respectively selected and based on the patient's previous answer, thus enabling precise measurement at individual level with few items. The working mechanism of a CAT can be illustrated by the following example: if a patient disagreed with the question "Are you too tired to take exercise?", this patient would not also get a question reflecting more severe fatigue as "Are you too tired to go for a short walk?".

For the development of a CAT for fatigue in RA, adequate items have to be selected for inclusion in an item bank.

Items of the aforementioned fatigue questionnaires (7-10) are available in Dutch. However, the patients' perspective was not part of their development and they were not specifically designed for patients with RA. Therefore, it is of particular importance to check content validity.

We conducted a Delphi study to allow patients, rheumatologists and nurses to evaluate fatigue items. It is essential to include the opinion of patients and professionals in the development of questionnaire items to ensure content validity (13). Only patients can report on the subjective experience of fatigue, and clinicians have the most experience with the outward manifestation of a symptom or condition (13).

This paper focuses on the evaluation of existing fatigue items: how do experts evaluate the items from traditional fatigue questionnaires (SF-36 subscale vitality, FACIT-F, POMS subscale fatigue/inertia, MAF)? Do the items of the BRAF-MDQ meet the meaning of fatigue in the Netherlands?

Patients and methods

Delphi process

With the Delphi process (14), opinions about a certain topic can be collected by a questionnaire, which is sent by (electronic) mail to a panel of potential participants who then fill it out individually. In our case, they rated the fatigue items on their own. Contrary to group discussion, no attention should be paid to prevent especially the more dominant participants from expressing their opinion, so that the phenomenon of "group think" cannot obstruct the viewing of different opinions. A Delphi study consists of at least two rounds and aims to reach consensus among the participants. After each Delphi round, systematic feedback on the results of the previous round are provided to the participants (15). By doing so, the experts are informed about the opinions that are present in the group. The extent of agreement is determined by statistical measures as no in-person meetings are conducted.

Preparation of the item pool

In our aim to develop a CAT for fatigue in RA, we need a large and comprehensive item pool to cover all dimensions of the fatigue experience in RA. We included 40 items of the validated Dutch versions of the RAND SF-36 subscale vitality (16), FACIT-F (17), POMS subscale fatigue/inertia (18), and MAF (19), henceforth referred to as traditional questionnaires. We also included all 20 items of the BRAF-MDQ, which we translated from English into Dutch using the recommended cross-cultural translation procedure (20). Additional items developed from interview material were enclosed (not reported in this paper, the entire process of the Delphi study is described elsewhere [21]).

Participants and data collection

We contacted the experts (40 rheumatologists, 40 nurses and 31 patients) by e-mail, informing them about the study and requesting their participation. The e-mails contained a link to an online questionnaire in which we presented the fatigue items. The e-mail addresses of patients were collated from the database of the Patient Research Partners of the Arthritis Centre Twente and those of the rheumatologists and nurses from the member list of a Dutch professional association in Rheumatology (NVR) and the DREAM registry. The participants were asked to indicate on a 4-point Likert scale (very appropriate, rather appropriate, less appropriate, not at all appropriate) how appropriate they thought each item was. The participants also had the possibility to comment on the items and suggest improvements in the blank fields below each item.

After each round, the data were analysed and summarised. For the items which had to be re-evaluated in round 2, information about its evaluation in round 1 was provided by presenting the relevant item, a brief summary of the comments given, and finally the reworded item. Items had to be evaluated in the same way as in round 1.

Analysis and criteria for item selection

In Delphi studies, different rules for defining sufficient consensus are applied (15), ranging from 55–80%. To establish the rule for our study, we searched the literature for other Delphi studies in health research. Repeatedly, the criterion of 80% was used for the selection of an item or topic, as for EULAR recommendations (22). On the basis of these examples, we also applied the conservative criterion of 80% agreement. Further properties of our rule were developed in discussion among the authors to meet our special situation with different expert groups.

An item was directly selected for inclusion in the CAT item pool if 80% or more of the participants rated it as appropriate (response options 'very appropriate' and 'rather appropriate'). Agreement by 50% or less of the participants led to the item's rejection.

Besides the global percentage, we also calculated the percentages per expert group. If the mean percentage was sufficiently high (80%) or low (50%), but one group differed by more than 10% from the criterion (below 70% or above 60%), the item was not directly selected or removed. They, and also those items rated as appropriate by 50%–80% of the experts, were discussed by the authors and re-worded according to the participants' comments and suggestions. If the same comment was raised by at least two participants, it led to an adaptation of the item. The re-worded items were judged again in round 2 by the panel. After round 2, comments were discussed among the researchers and the final decision about in- or exclusion of the items was made.

Results

In the first round, 15 patients, 10 rheumatologists and 20 nurses, working at 21 hospitals spread over the Netherlands, returned the completed questionnaire (n=45). In the second round, 80% of the participants of the first round (15 nurses, 13 patients and 8 rheumatologists) participated (n=36). In both rounds, our expert panel clearly consisted of more women than men (2:1) and the mean age was 51.4 (SD=11.4) in the first round and 52.4 (SD=10.5) in the second round.

Evaluation of the items

Round 1. An immediate decision could be made on 18 of the 40 items of the traditional questionnaires (SF-36 subscale vitality, FACIT-F, POMS subscale fatigue/inertia, MAF). Sixteen items (40%) were rated as appropriate by at least 80% of the participants so that they were directly selected for the development of the CAT. Two items were rejected because they were rated as appropriate by 50% or less of the participants. Item 7 of the POMS "Over the past 7 days I felt bushed" was excluded. Participants rated this item as too negative, dramatic, extreme or exaggerated. Moreover, they noted that the item might be confusing because it could also reflect physical circumstances such as a loss of strength in the joints, or have a psychological component such as depressive feelings. The second excluded item was question 2 of the MAF "How severe is the fatigue which you have been experiencing during the past week? 1=mild 2 3 4 5 6 7 8 9 10=severe". In the official Dutch questionnaire, fatigue was translated as '*uitputting*' which actually means 'exhaustion' and it was criticised that exhaustion is always assumed to be present. Furthermore, participants said that it was not possible to measure exhaustion with the used scale since you are either exhausted or not, *i.e*. that exhaustion is always serious.

Regarding the BRAF-MDQ, 12 out of the 20 items (60%) were directly selected and 1 item (5%) was excluded. This was question 12 "Over the past 7 days, have you felt embarrassed because of fatigue?", which participants found ambiguous and too difficult.

The authors discussed the remaining items (>50% and <80% experts agreed) of the five scales, whereby the participants' comments and suggestions were considered. Items that did not receive comments on their content but on their formulation or the response options, for example, were adjusted (phrase or word replaced, deleted or added) and included in the item pool. This applied to 17 (42.5%) items of the traditional questionnaires. A major concern was again the use of the Dutch word for exhaustion in the MAF. Participants found this term inappropriate and advised using another word that reflects

the word 'fatigue' better. Accordingly, we replaced the term '*uitputting*/exhaustion' with '*vermoeidheid*/fatigue' in all concerned MAF items and, as respondents found the formulation of most MAF items cumbersome, we shortened them. Furthermore, experts rated the response options of the FACIT-F as unclear, so we re-worded these too. Table I shows examples of these adaptations. Item 5 of the POMS "Over the past 7 days I felt sluggish" was excluded after discussion, the reason being that respondents rated this item as unclear and no alternatives were found for reformulation. The authors selected 3 (15%) BRAF-MDQ items after minor adaptations, such as the replacement or deletion of a word.

Four items (10%) of the traditional questionnaires needed more extensive adaptations (*e.g.* inclusion of examples, reformulation of an entire item) and were presented for re-evaluation in round 2. Four BRAF-MDQ items (20%) were presented for re-evaluation either because examples had to be inserted or because only criticism but no suggestion for improvement had been given.

Round 2. Table II shows examples of items included in round 2. After the second round, 3 adapted items of the traditional questionnaires were evaluated as adequate by more than 80% of the participants and could immediately be selected. Participants had no complaints about the general adaptations made to these items (*e.g.* different word in MAF, response options FACIT-F). Of the 4 BRAF-MDQ items included in round 2, two were rated as adequate by at least 80% of the participants and immediately selected.

Based on the comments given, the authors decided to exclude one item of the traditional questionnaires, namely question 1 of the POMS (see Table II). This item was adjusted according to the comments; the Dutch translation for 'worn out' (*doodop*) was replaced with a synonym (*bekaf*) and the response options were adapted. Despite this, many participants still found this item unclear and too negative. Two BRAF-MDQ items were included with adapted response after discussion between the authors.

Table I. Examples for adaptations and selection of items by the authors.

Item in round 1	Comments	Adapted and selected item
MAF item 5 In the past week, to what degree has fatigue interfered with your ability to cook? 1=not at all 2 3 4 5 6 7 8 9 10=a great deal	Formulate item and response options easier, replace the Dutch word for exhaustion (*uitputting*) with a word that refers to fatigue better (*vermoeidheid*) because exhaustion is too extreme, it is possible that cooking is not applicable to a patient.	During the past 7 days, how did your fatigue prevent you from cooking? 1=not at all, 2=a little, 3=rather, 4=to a great extent, 5=does not apply to me
FACIT-F item 4 During the last 7 days, I felt tired. 0=not at all, 1=a little bit, 2=somewhat, 3=quite a bit, 4=very much	The response options are unclear (translated in Dutch as 0=*helemaal niet*, 1=*een beetje*, 2=*enigzins*, 3=*in vrij hoge mate*, 4=*in zeer hoge mate*); the formulation of option 3 and 4 is too formal, and the difference between option 1 and 2 is not clear.	During the past 7 days, I felt tired. 1=not at all, 2=a little, 3=rather, 4=to a great extent (Dutch: 1=*helemaal niet*, 2=*een beetje*, 3=*nogal*, 4=*in sterke mate*)

Table II. Examples of adapted items presented for re-evaluation in round 2.

Item origin	Original item	Adapted item
Finally selected		
FACIT-F item 2	During the last 7 days I felt weak all over. 0=not at all, 1=a little bit, 2=somewhat, 3=quite a bit, 4=very much	During the past 7 days, my fatigue made me feel weak. 1=not at all, 2=a little, 3=rather, 4=to a great extent
MAF item 3	In the past week to what degree has fatigue caused you distress? 1–10/1=no distress, 10=a great deal of distress	Was your fatigue during the past 7 days aggravating? 1=not at all, 2=a little, 3=rather, 4=to a great extent
MAF item 14	In the past week, to what degree has fatigue interfered with your ability to exercise, other than walking? 1=not at all 2 3 4 5 6 7 8 9 10=a great deal	During the past 7 days, how did your fatigue prevent you from moving (*e.g.* walking, cycling, playing sports)? 1=not at all, 2=a little, 3=rather, 4=to a great extent
BRAF-MDQ item 12	Have you lacked mental energy because of fatigue? not at all, a little, quite a bit, very much	Have you lacked mental energy (*e.g.* energy to think properly about sth., to make plans, to be creative) because of fatigue? not at all, a little, quite a bit, very much
BRAF-MDQ item 20	Have you felt down or depressed because of fatigue? not at all, a little, quite a bit, very much	Have you felt down or dejected because of fatigue? not at all, a little, quite a bit, very much
Finally excluded		
POMS item 1	Over the past 7 days I felt worn out (Dutch translation: *doodop*). 1=not at all, 2=a little, 3=moderate, 4=quite a bit, 5=extremely	Over the past 7 days I felt worn out (Dutch: *bekaf*). 1=never, 2=rarely, 3= sometimes, 4=usually, 5=always

Table III. Selection of fatigue items from the four traditional scales and the BRAF-MDQ.

Scale/ Evaluation	RAND-36 4 items	FACIT-F 13 items	POMS 7 items	MAF 16 items	BRAF-MDQ 20 items	Summary 60 items
Round 1						
Directly selected	3 (75%)	9 (69.2%)	1 (14.3%)	3 (18.8%)	12 (60%)	28 (46.6%)
Selected after adaptation	1 (25%)	3 (23.1%)	3 (42.9%)	10 (62.5%)	3 (15%)	20 (33.3%)
Directly rejected			1 (14.3%)	1 (6.25%)	1 (5%)	3 (5%)
Rejected after discussion			1 (14.3%)			1 (1.7%)
Adapted and selected for re-evaluation in round 2	0	1	1	2	4	8
Round 2						
Directly selected		1 (7.7%)		2 (12.5%)	2 (10%)	5 (8.3%)
Selected after discussion					2 (10%)	2 (3.3%)
Directly rejected						
Rejected after discussion			1 (14.3%)			1 (1.7%)
Finally **excluded** items	0	0	3	1	1	5 (8%)

Finally, 36 items (90%) from the traditional questionnaires and 19 items (95%) from the BRAF-MDQ were included in our item pool. Table III shows the percentage of items per scale in each selection step.

Discussion

This Delphi study provided insights into the perspective of patients and professionals on fatigue items of four traditional scales (SF-36 subscale vitality, FACIT-F, POMS subscale fatigue/inertia and MAF) and a new questionnaire (BRAF-MDQ) developed in the UK.

It appeared that only 40% of the items from traditional questionnaires were clear and adequate in their original style for at least 80% of the participants. This finding underlines the relevance of the perspective of professionals and patients in the development of measurement instruments for fatigue in RA. However, most of the items could be selected after adaptations, with only 4 of the 40 items being excluded. Remarkably, three of these items originated from the POMS. One item was rated as too extreme and confusing, because it could also be related to circumstances other than fatigue; the other items were also rated as too extreme or as unclear. The fourth excluded item came from the MAF and had to be directly excluded due to insufficient percentages of ratings as adequate. The item would probably have received more positive evaluations, though, with the word '*vermoeidheid*' that we used to replace '*uitputting*' in all of the selected MAF items. The reaction of our experts to the term '*uitputting*' suggests that the Dutch version of the MAF might sooner measure exhaustion than fatigue. This finding points to the need to carefully check and re-check cross-cultural translations to ensure that the intended measurement aim is met. However, our results cannot be generalised to the items in the original English questionnaire since it was the official Dutch translations that were evaluated in this study. An evaluation of the original English items by rheumatologists, nurses and patients could provide more insight.

The BRAF-MDQ turned out to suit the fatigue experience in a Dutch population well. This may be attributed to its advantage of including the patients' perspective in the development process. Only one item had to be excluded and was about feeling embarrassed. It is difficult to fully explain why it received low percentages of agreement since only few comments were given. A possible explanation might be that British patients might feel more embarrassed about fatigue than Dutch patients since they live in a so-called 'stiff upper lip', masculine culture in which it is not so common to show weakness (2, 23). In contrast, the Netherlands is considered a more feminine and openly nurturing culture (23).

This study was limited to the Netherlands, so we are not able to draw conclusions about the quality of the examined items in other countries. Furthermore, there were no generally accepted criteria we could apply for our in- and exclusion of items. However, we carefully deliberated our criteria and chose for the conservative criterion of 80% agreement as inclusion criterion that was frequently used in other Delphi studies (22). A considerable amount of the criticism on the fatigue items was related to translation difficulties rather than to their content. To conclude, 90% of the traditional questionnaire items and 95% of the BRAF-MDQ items showed good content validity and feasibility (partly in an adapted version) for use in the Netherlands and will be included in our item pool for the development of a CAT for fatigue in RA. This item pool contains further items that were validated by our expert panel, for example items based on interview material (21). We assume that our item pool is comprehensive and adequate for the measurement of fatigue in RA. In a next step, the items have to be scaled according to item response theory (IRT). With IRT, item parameters as the difficulty level can be assessed for each item independently (24). This information is required to ideally match the items to the patient's individual level and construct the CAT for fatigue in RA.

Acknowledgements

We would like to thank the following participants in this study: patients from

the forum of Patient Research Partners of the Arthritis Centre Twente, and rheumatologists and nurses from Universitair Medisch Centrum Groningen, Medisch Centrum Leeuwarden, Ziekenhuisgroep Twente, Zorg Groep Twente, Medisch Spectrum Twente, St. Maartenskliniek Nijmegen, TweeSteden Ziekenhuis Locatie Tilburg, Universitair Medisch Centrum Utrecht, Rijnstate Ziekenhuis Arnhem, Meander Medisch Centrum Amersfoort, Ziekenhuis Gelderse Vallei Ede, Medisch Centrum Haaglanden s-Gravenhage, Tweesteden Ziekenhuis Tilburg, Evean Thuiszorg Purmerend, Streekziekenhuis Midden Twente Hengelo, Isala Klinieken Zwolle, Albert Schweitzer Ziekenhuis Dordrecht, Spaarne Ziekenhuis Hoofddorp, Scheperziekenhuis Emmen, Thuiszorg West-Brabant Roosendal, Omring Thuiszorg Hoorn.

References

1. MINNOCK P, KIRWAN J, BRESNIHAN B: Fatigue is a reliable, sensitive and unique outcome measure in rheumatoid arthritis. *Rheumatology* 2009; 48: 1533-6.
2. HEWLETT S, COCKSHOTT Z, BYRON M *et al*.: Patients' perceptions of fatigue in rheumatoid arthritis: overwhelming, uncontrollable, ignored. *Arthritis Care Res* 2005; 53: 697-702.
3. NIKOLAUS S, BODE C, TAAL E, VAN DE LAAR MAFJ: Four different patterns of fatigue in rheumatoid arthritis patients: results of a Q-sort study. *Rheumatology* 2010; 49: 2191-9.
4. VAN OERS ML, BOSSEMA ER, THOOLEN BJ *et al*.: Variability of fatigue during the day in patients with primary Sjögren's syndrome, systemic lupus erythematosus, and rheumatoid arthritis. *Clin Exp Rheumatol* 2010; 28: 715-21.
5. STRAND V, KHANNA D: The impact of rheumatoid arthritis and treatment on patients' lives. *Clin Exp Rheumatol* 2010; 28: S32-40.
6. HEWLETT S, HEHIR M, KIRWAN JR: Measuring fatigue in rheumatoid arthritis: a systematic review of scales in use. *Arthritis Rheum* 2007; 57: 429-39.
7. WARE JE JR, SHERBOURNE CD: The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; 30: 473-83.
8. CELLA D, YOUNT S, SORENSEN M, CHARTASH E, SENGUPTA N, GROBER J: Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. *J Rheumatol* 2005; 32: 811-9.
9. MCNAIR D, LORR M, DROPPELMAN L: Profile of Mood States Manual. New York: Multi-health Systems Inc; 1992.
10. BELZA TACK B: Dimensions and correlates of fatigue in older adults with rheumatoid arthritis [dissertation]. San Francisco: University of California; 1991.
11. NICKLIN J, CRAMP F, KIRWAN J, GREENWOOD R, URBAN M, HEWLETT S: Measuring fatigue in rheumatoid arthritis: A cross-sectional study to evaluate the Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional questionnaire, visual analogue scales, and numerical rating scales. *Arthritis Care Res* 2010; 62: 1559-68.
12. ROSE M, BEZJAK A: Logistics of collecting patient-reported outcomes (PROs) in clinical practice: an overview and practical examples. *Qual Life Res* 2009; 18: 125-36.
13. STREINER DL, NORMAN GR (Eds.): Health measurement scales – a practical guide to their development and use. New York: Oxford University Press; 2003.
14. JONES J, HUNTER D: Qualitative research: consensus methods for medical and health services research. *BMJ* 1995; 311: 376-80.
15. Powell C: The Delphi technique: myths and realities. *J Adv Nurs* 2003; 41: 376-82.
16. VAN DER ZEE KI, SANDERMAN R: Het meten van de algemene gezondheidstoestand met de RAND-36 – een handleiding. Noordelijk Centrum voor Gezondheidsvraagstukken, Rijksuniversiteit Groningen. 1993.
17. FACIT FUNCTIONAL ASSESSMENT OF CHRONIC ILLNESS THERAPY: Retrieved July 16, 2009, from www.facit.org
18. CLUYDTS RJG: Gemoedstoestanden en slaap – een experimenteel onderzoek [dissertation]. Brussel: Vrije Universiteit Brussel; 1979.
19. MAPI INSTITUTE: Retrieved July 31, 2009, from www.mapi-institute.com
20. GUILLEMIN F, BOMBARDIER C, BEATON D: Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993; 46: 1417-32.
21. NIKOLAUS S, BODE C, TAAL E, VAN DE LAAR MAFJ: Selection of items for a computer-adaptive test to measure fatigue in patients with rheumatoid arthritis – A Delphi approach. Qual Life Res 2011 Epub July 31.
22. AVOUAC J, KOWAL-BIELECKA, O, LANDEWÉ R *et al*.: European League Against Rheumatism (EULAR) Scleroderma Trial and Research group (EUSTAR) recommendations for the treatment of systemic sclerosis: methods of elaboration and results of systematic literature research. *Annals of the Rheumatic Diseases* 2009; 68: 629-34.
23. GEERT HOFSTEDE™ CULTURAL DIMENSIONS: Retrieved July 22, 2010, from www.geert-hofstede.com/hofstede_dimensions.php?culture1=62&culture2=94
24. HAMBLETON RK, SWAMINATHAN H & ROGERS HJ: Fundamentals of Item Response Theory. Sage Publications. (1991).