BMC Medical Research Methodology

Research article

Performance of health-status scales when used selectively or within multi-scale questionnaire

Christina Gummesson*[1], Isam Atroshi[2] and Charlotte Ekdahl[1]

Address: [1]Department of Physical Therapy, Lund University, Sweden and [2]Department of Orthopedics, Hässleholm-Kristianstad Hospitals, Kristianstad, Sweden

Email: Christina Gummesson* - christina.gummesson@sjukgym.lu.se; Isam Atroshi - isam.atroshi@mailbox.swipnet.se; Charlotte Ekdahl - charlotte.ekdahl@sjukgym.lu.se

* Corresponding author

Abstract

Background: Little work has been done to investigate the suggestion that the use of selected scales from a multi-scale health-status questionnaire would compromise reliability and validity. The aim of this study was to compare the performance of three scales selected from the SF-36 generic health questionnaire when administered in isolation or within the entire SF-36 to patients with musculoskeletal disorders.

Methods: Two groups of patients referred to an orthopedic department completed a mailed questionnaire within 4 weeks prior to and a second questionnaire during their visit. The first group completed three SF-36 scales related to physical health (physical functioning, bodily pain, and general health perceptions) on one occasion and all eight SF-36 scales on the other occasion. The second group completed the entire SF-36 on two occasions. Results for patients who reported unchanged health status and had complete scores were analyzed; 80 patients in the first and 62 patients in the second group.

Results: The Cronbach alpha reliability and intraclass correlation coefficients exceeded 0.7 for all three scales for both groups. For the first group the mean difference between the scores was 0.4 point for physical functioning, 2.5 points for bodily pain, and 0.5 point for general health perceptions, which did not differ significantly from the corresponding differences for the second group (0.1, 1.9 and 1 point, respectively).

Conclusion: The use of selected scales from a multi-scale health-status questionnaire seems to yield similar results compared to their use within the entire questionnaire.

Background

Measures of health status and quality of life are being increasingly used in clinical research. In the evaluation of many conditions, it might be necessary to combine generic and disease-specific questionnaires. Many questionnaires are long and consist of several scales, which might substantially increase responder burden. Generic health-status questionnaires usually consist of separate scales related to physical and mental health. In musculoskeletal conditions, physical health scales are more likely to show differences after treatments and would thus be used in sample size estimations; mental health scales would then lack the power to show differences. It has been suggested that multi-scale health-status questionnaires should be used in their entirety and that the use of selected scales

would, by taking them out of their context, compromise their reliability and validity and the possibility to compare scores across studies and with population norms [1]. However there is little scientific work concerning the influence of excluding some scales in a health-status questionnaire on the performance of the remaining scales. Demonstrating whether the scores yielded when using selected scales are similar to those yielded when the entire questionnaire is administered would be important because similarity of scores would allow comparison with the corresponding scores in studies that used the entire questionnaire and with population norms. This would facilitate the interpretation of scores when selected scales are used.

The SF-36 is a widely used health-status measure that consists of eight scales related to physical and mental health [2–4]. Different SF-36 scales have been used selectively in previous studies without prior evidence of reliability and validity [5–7]. The purpose of this study was to investigate the performance of three SF-36 scales related to physical health (physical functioning, bodily pain and general health perceptions) when administered selectively or within the entire SF-36 to a patient population with musculoskeletal disorders.

Methods
This 2-part study was conducted on patients with musculoskeletal disorders referred from primary care physicians to the only orthopedic department available in the study region. All referred patients, aged 25 to 74 years, who had a scheduled visit to the orthopedic department during a 6-week period were asked to complete a mailed questionnaire within 4 weeks before their visit and to complete a second questionnaire administered during the visit.

Consecutive administration of selected scales and entire questionnaire
In the first part of the study, one questionnaire comprised three SF-36 scales related to physical health (physical functioning, bodily pain, and general health perceptions) without any modifications in the order or composition of the items. The second questionnaire comprised all eight SF-36 scales with no modifications. During the first half of the study period the first questionnaire comprised the three selected SF-36 scales and the second questionnaire the entire SF-36; in the second half of the study period the two questionnaires were administered in reverse order. On both occasions the questionnaires were self-completed by the patients.

Repeated administration of entire questionnaire
In the second part of the study a formal test-retest reliability assessment of the SF-36 was performed; the entire SF-36 was administered on two occasions in a similar fashion as in the first part of the study.

Item concerning change in health status
In both groups the questionnaire that was administered on the second occasion started with an inquiry about current health status compared to that when the first questionnaire was completed (Question: Compared to when you completed the questionnaire regarding your health about a week ago, how is your health now? response options; much better, somewhat better, same, somewhat worse, much worse).

Statistical analysis
The reliability (internal consistency) of the SF-36 physical functioning, bodily pain and general health perceptions scales was assessed with the Cronbach alpha coefficient [8]. The item scores for each scale were transformed into scale scores ranging from 0 (worst) to 100 (best) [1]. The mean score and 95% confidence interval (CI) for each of the three scales were calculated. The agreement between the scores for each of the three scales administered as isolated scales and within the entire SF-36 was assessed using the intraclass correlation coefficient (ICC) and the differences were tested with the paired t-test [8]. This analysis included only the patients who reported unchanged health status at the time of completing the second questionnaire. Only questionnaires with complete responses for all items in all of the three scales were included in the analysis. Because the analysis involved assessment of agreement missing data were not replaced. The same analyses were performed on the data obtained when the entire SF-36 was administered on two occasions. The mean differences between the scores shown when the three scales were selectively administered and those shown when they were administered within the entire SF-36 were compared to the mean differences in the scores shown after repeated administration of the entire SF-36 using the t-test.

Results
Consecutive administration of selected scales and entire questionnaire
During the 6-week study period, 137 consecutive referred patients attended the orthopedic department for a scheduled visit. Of these, 11 completed only one of the questionnaires, and 23 reported changed health status since completing the first questionnaire. The remaining 103 patients completed both questionnaires and reported unchanged health status. For 23 (22%) of these patients scores could not be computed for at least one scale because of missing item responses. Thus, 80 patients had scores for all three scales for both occasions. The mean age of these 80 patients was 50 (SD, 11) years and 41 (51%) were women. The mean time interval between the responses to the two questionnaires was 14 (SD, 3) days.

Table 1: Reliability of the physical functioning, bodily pain, and general health perceptions scales when administered independently on one occasion (A) and within the entire SF-36 on a second occasion (B)

Scale (n = 80)	Alpha[a]		ICC (95% CI)[b]	Mean (SD)		Mean difference (95% CI)
	A	B		A	B	
Physical functioning	0.89	0.88	0.93 (0.90–0.96)	64.3 (24)	64.8 (24)	0.4 (-1.5–2.4)
Bodily pain	0.87	0.81	0.78 (0.68–0.86)	36.8 (18)	39.4 (20)	2.5 (-.0.3–5.4)
General health perceptions	0.86	0.85	0.87 (0.80–0.91)	58.6 (23)	59.1 (23)	0.5 (-2.2–3.2)

[a] Cronbach alpha coefficient values range from 0 (no) to 1 (perfect) reliability. [b] Intraclass correlation coefficient (ICC) values range from 0 (no) to 1 (perfect) reliability. CI, confidence interval

The Cronbach alpha reliability coefficient exceeded 0.8 for all three scales (Table 1). The ICC was good for all three scales and the mean difference between the scores was 0.4 point for the physical functioning scale, 2.5 points for the bodily pain scale, and 0.5 point for the general health perceptions scale, indicating good agreement between the scores when the three scales were administered with and without the remaining SF-36 scales.

Repeated administration of entire questionnaire
In the second part of the study, 107 consecutive referred patients attended their scheduled visit during a 6-week period. Of these, 18 completed only one of the questionnaires, and 15 reported changed health status since completing the first questionnaire. The remaining 74 patients completed both questionnaires and reported unchanged health status. For 12 (16%) of these patients scores could not be computed for at least one of the three scales studied because of missing item responses. Thus, 62 patients had scores for all three scales for both occasions. The mean age of these 62 patients was 51 (SD, 11) years and 34 (55%) were women. The mean time interval between the responses to the two questionnaires was 13 (SD, 5) days. The Cronbach alpha reliability coefficient exceeded 0.7 for all three scales (Table 2). The ICC was good for all three scales and the mean difference between the scores was 0.1 point for the physical functioning scale, 1.9 points for the bodily pain scale, and 1 point for the general health perceptions scale, indicating good test-retest reliability.

Comparison of score differences
For all three scales, the mean differences between the scores obtained when the three scales were selectively administered and those obtained when they were administered within the entire SF-36 did not differ significantly from the mean differences shown after repeated administration of the entire SF-36. The mean difference (95% CI)

for the physical functioning scale was 0.3 (-2.8–3.4), for the bodily pain scale 0.6 (-3.5–4.7), and for the general health perceptions scale -0.5 (-4.4–3.4).

Discussion
This study showed that the physical functioning and general health perceptions scales gave similar scores when administered independently or within the entire SF-36. Although the bodily pain scale showed a difference of 2.5 points, this occurred in a patient population with musculoskeletal disorders causing pain, the severity of which was rated on two occasions. A difference of approximately 2 points also was found when the entire SF-36 was administered on two occasions. Although no test-retest reliability data have been presented for most of the published SF-36 population norms, one study performed on patients with rheumatoid arthritis showed an intraclass correlation coefficient for the physical functioning, bodily pain, and general health perceptions scales of 0.93, 0.76, and 0.91, and mean score difference of -1.8, 2.9 and 0.2, respectively [9].

The findings of the present study do not support the suggestion that the exclusion of some scales of a health-status measure would influence the response patterns to the remaining scales. We have not found any previous study on the influence of excluding some scales in a health-status questionnaire. Specific diseases might have substantial impact on certain health dimensions and little or no impact on others, which would be reflected on the scores for the scales measuring these dimensions. Also, health-status scales are often used as part of more extensive questionnaires and researchers might elect to include selected scales that are relevant to the study purpose; the physical functioning, bodily pain and general health perceptions scales have been selectively used previously [5–7]. Shorter versions of certain health-status questionnaires have been introduced with the purpose of reducing responder bur-

Table 2: Reliability of the physical functioning, bodily pain, and general health perceptions scales when the entire SF-36 was administered on two occasions (T1 & T2)

Scale (n = 62)	Alpha[a]		ICC (95% CI)[b]	Mean (SD)		Mean difference (95% CI)
	T1	T2		T1	T2	
Physical functioning	0.88	0.88	0.92 (0.86–0.95)	67.3 (23)	67.4 (23)	0.1 (-2.4–2.6)
Bodily pain	0.86	0.89	0.86 (0.78–0.92)	44.7 (21)	46.7 (23)	1.9 (-1.0–4.9)
General health perceptions	0.79	0.72	0.83 (0.73–0.89)	61.0 (20)	62.0 (17)	1.0 (-1.8–3.8)

[a] Cronbach alpha coefficient values range from 0 (no) to 1 (perfect) reliability. [b] Intraclass correlation coefficient (ICC) values range from 0 (no) to 1 (perfect) reliability. CI, confidence interval

den. However, these shorter versions attenuate the original scales and might not perform as well in diseases that have larger impact on specific scales. The SF-12 (a shorter version of the SF-36) generates a physical and a mental component summary score [10]. These summary scales have demonstrated inferior performance compared to the bodily pain or physical functioning scales in musculoskeletal conditions [11]. Shorter questionnaires might have a higher response rate (although not consistently shown) [12,13], in addition to saving time and resources. By reducing the workload required, shorter questionnaires might facilitate the participation of clinicians in national databases improving the validity of the information obtained from these databases. Use of existing scales might be an alternative to the long process of constructing shorter questionnaires followed by extensive reliability and validity testing [14].

However, excluding certain scales has been discouraged by the questionnaire's developers on the basis that it might compromise reliability [1]. A previous study evaluated the use of selected scales but examined only reliability of these scales without showing whether they would generate similar scores when used within the entire questionnaire [15]. Demonstrating good psychometric properties of selectively used scales is important. However, maintaining good reliability of selected SF-36 scales does not necessarily ensure that the scales would yield similar scores as when administered within the entire SF-36 to allow comparison across studies. The findings of the present study imply that the scores of the selectively used scales can be compared with the corresponding scores in studies that used the entire SF-36 and with population norms, thus facilitating score interpretability.

In the first part of the study, the order in which the two questionnaires were administered was not random. However, it is unlikely that this could have influenced the re-

sults because consecutive patients were included and the questionnaires were self-completed.

Conclusions
The use of selected scales from a multi-scale health-status questionnaire seems to yield similar results compared to their use within the entire questionnaire.

Competing interests
None

Authors' contributions
CG and IA participated in the design of the study, data collection and analysis, and writing of this manuscript. CE participated in the analysis and writing of this manuscript. All authors read and approved the final manuscript.

Acknowledgments

References
1. Ware JE, Snow KK, Kosinski M and Gandek B **SF-36 health survey manual and interpretation guide.** *Boston: New England Medical Center* 1993,
2. Ware JE and Sherbourne CD **The MOS 36-item short-form health survey (SF-36): conceptual framework and item selection.** *Med Care* 1992, **30**:473-483
3. Jenkinson C, Coulter A and Wright L **Short form 36 (SF36) health survey questionnaire: normative data for adults of working age.** *BMJ* 1993, **306**:1437-1440
4. Sullivan M, Karlsson J and Ware JE **The Swedish SF-36 health survey: evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden.** *Soc Sci Med* 1995, **41**:1349-1358
5. Gottschalk A, Smith DS, Jobes DR, Kennedy SK, Lally SE, Noble VE, Grugan KF, Seifert HA, Cheung A, Malkowicz SB, Gutsche BB and Wein AJ **Preemptive epidural analgesia and recovery from radical prostatectomy: a randomized controlled trial.** *JAMA* 1998, **279**:1076-1082
6. Unwin C, Blatchley N, Coker W, Ferry S, Hotopf M, Hull L, Ismail K, Palmer I, David A and Wessely S **Health of UK servicemen who served in Persian Gulf War.** *Lancet* 1999, **353**:169-178

7.   Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, Hollingsworth JC, Ashton CM and Wray NP **A controlled trial of arthroscopic surgery for osteoarthritis of the knee.** *N Engl J Med* 2002, **347:**81-88

8.   Streiner DL and Norman GR **Health measurement scales: a practical guide to their development and use.** *Oxford: Oxford University Press* 1995,

9.   Ruta DA, Hurst NP, Kind P, Hunter M and Stubbings A **Measuring health status in British patients with rheumatoid arthritis: reliability, validity and responsiveness of the short form 36-item health survey (SF-36).** *Br J Rheumatol* 1998, **37:**425-436

10. Ware J, Kosinski M and Keller SD **A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity.** *Med Care* 1996, **34:**220-233

11. Atroshi I, Gummesson C, Johnsson R and Sprinchorn A **Symptoms, disability, and quality of life in patients with carpal tunnel syndrome.** *J Hand Surg [Am]* 1999, **24:**398-404

12. Kalantar JS and Talley NJ **The effects of lottery incentive and length of questionnaire on health survey response rates: a randomized study.** *J Clin Epidemiol* 1999, **52:**1117-1122

13. Eaker S, Bergstrom R, Bergstrom A, Adami HO and Nyren O **Response rate to mailed epidemiologic questionnaires: a population-based randomized trial of variations in design and mailing routines.** *Am J Epidemiol* 1998, **147:**74-82

14. Cook DJ, Guyatt GH, Juniper E, Griffith L, McIlroy W, Willan A, Jaeschke R and Epstein R **Interviewer versus self-administered questionnaires in developing a disease-specific, health-related quality of life instrument for asthma.** *J Clin Epidemiol* 1993, **46:**529-534

15. Raina P, Bonnett B, Waltner-Toews D, Woodward C and Abernathy T **How reliable are selected scales from population-based health surveys? An analysis among seniors.** *Can J Public Health* 1999, **90:**60-64

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2288/3/3/prepub