

MATTERS ARISING

Should the MHAQ ever be used?

We read with interest the article by Stucki and colleagues,¹ and would like to make some comments in relation to similar work performed in our hospital.

In 1994 we gave a questionnaire to 122 rheumatoid ambulatory patients, including the disability section of the Health Assessment Questionnaire (HAQ) and questions regarding pain, global health, and functional status. When we compared the results using the 20 items of the original HAQ versus the eight extracted items corresponding to the modified HAQ (MHAQ), we observed significant differences in scores between both tests. The MHAQ total scores were on average 31.5% lower than those of the full HAQ (1.49 (SD 0.74) for the HAQ and 1.02 (1.00) for the MHAQ; $p < 0.001$ by paired t test). In addition, significant differences were found between all the subscales.

Internal consistency was high for both tests ($\alpha > 0.9$) and there was a good correlation between them, both for total scores (Pearson's $r = 0.88$) and for each of the subscales ($0.59 < r < 0.87$), that was highly significant ($p < 0.0001$). Correlations of HAQ and MHAQ with functional status, pain, and global health (measured using both visual analogue and Likert scales) were almost identical for both tests.

What conclusions can we draw from these data and previous work? First, the HAQ and the MHAQ score significantly differently, so we cannot merge patients or compare directly either studies or populations studied using these different tests. Though this effect was already present in the data of the original work by Pincus *et al.*,² those authors (unlike others^{1,3}) did not mention it specifically. We disagree with the assumption made by Stucki and colleagues that the difference found 'indicates that the HAQ contains more difficult items than the MHAQ', but we believe, rather, that this is the result of the specific scoring system used here. Both in the HAQ and in the MHAQ, each subscale is rated taking the highest value of the items composing the subscale. As the HAQ has several items in each subscale and the MHAQ only one, this means two to four opportunities more for the HAQ to score worse for each subscale compared with the MHAQ. In other words, the greater number of items in the HAQ gives it more chances to score higher than the MHAQ, and this is the most likely explanation for the differences found. This same reason serves to explain why the HAQ is more sensitive to change than the MHAQ.

An additional point, usually overlooked, is that the original HAQ has several complementary questions regarding the use of instruments or other people's aid to help in performing the activities that form the eight subscales. In the event that any one of the questions receives a positive response, the corresponding subscale is automatically scored at least as 2 or 'very difficult' to perform, even if the values of its particular items are lower; if they are higher, then the scale is scored as

3 or 'impossible' to perform. These complementary questions were not included in the MHAQ, which may also help to explain its lower scores compared with the HAQ. Furthermore, the use of these additional questions is not always made clear in the methods section of work published on this subject, and we feel that many researchers are obviating them. It would be interesting to study the effect on the final score of the HAQ of suppressing these questions, as this test would be easier to fill in and score without them.

After all these considerations, few arguments remain in support of the use of the MHAQ to assess functional capacity. Despite of the simplicity of the MHAQ, the HAQ can be completed by >80% of patients in about three minutes and scored in less than half that time,⁴ which is not long, even in the busiest clinics. The MHAQ is reliable and consistent, but the HAQ has the advantage of a greater sensitivity, better powers of discrimination, and much lower ceiling effects. In addition, a failure to respond to any of the questions of the MHAQ will lead to the relevant subscale being missing, while the HAQ has alternative items for each subscale. Moreover, the original purpose of the MHAQ was not as an alternative or 'short form' of the HAQ, but to allow for additional questions assessing patient satisfaction while keeping the questionnaire of a manageable size. We conclude that there is no reason at present to favour the use of the MHAQ for measuring functional capacity in clinical settings, and that the HAQ should be the preferred tool for this purpose in those rheumatic diseases with which it has proved useful.

MIGUEL A BELMONTE SERRANO
JUAN BELTRÁN FABREGAT
JAVIER OLMEDO GARZÓN
Unidad de Reumatología,
Hospital General de Castellón,
12003-Castellón, Spain

- 1 Stucki G, Stucki S, Brühlmann P, Michel B A. Ceiling effects of the Health Assessment Questionnaire and its modified version in some ambulatory rheumatoid arthritis patients. *Ann Rheum Dis* 1995; 54: 461-5.
- 2 Pincus T, Summey J A, Soraci S A Jr, Wallston K A, Hummon N P. Assessment of patient satisfaction in activities of daily living using a modified Stanford health assessment questionnaire. *Arthritis Rheum* 1983; 26: 1346-53.
- 3 Blalock S J, Sauter V H, DeVellis R F. The modified health assessment questionnaire difficulty scale. *Arthritis Care Res* 1990; 3: 182-8.
- 4 Wolfe F, Kleinhenkels S M, Cathey M A, Hawley D, Spitz P W, Fries J F. The critical value of the Stanford health assessment questionnaire functional disability index in patients with rheumatoid arthritis. *J Rheumatol* 1988; 15: 1480-8.

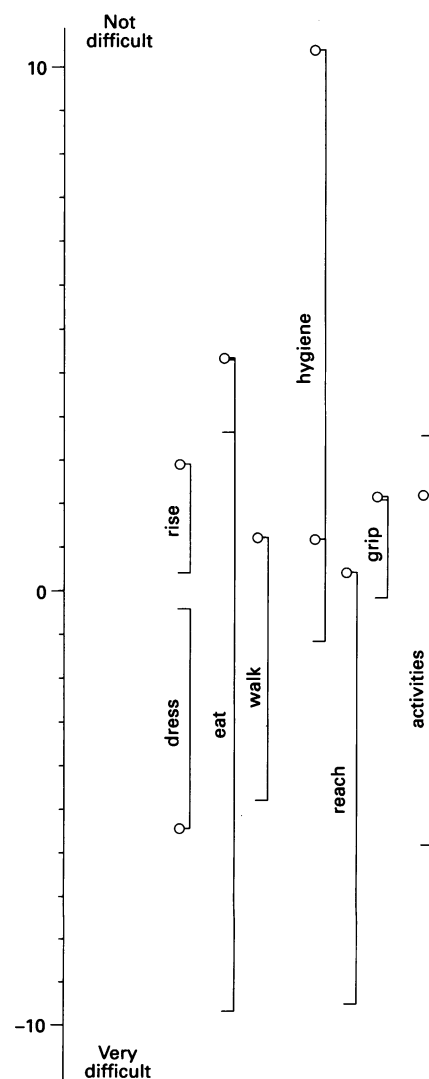
AUTHORS' REPLY:

The data presented by Drs Belmonte Serrano and Fabregat emphasise again two important aspects of the use of health status instruments. First, the MHAQ score is lower than the HAQ score, and thus HAQ and MHAQ scores are not interchangeable. Second, cross-sectional analysis is not sufficient when testing the metric properties of a short form compared with a full version. In cross-sectional analysis, the internal consistency of the short instrument may be as good as that of the long instrument, the correlation between the short and long version may be high, and the correlation with external parameters may be similar. However, as we showed in our study and as is emphasised by the data in the preceding letter, a short

version may not discriminate patients as well, and may fail to document clinically meaningful change as a result of a floor or ceiling effect.

There are indeed different reasons why the MHAQ score may be lower than the HAQ score and we agree that chance and grading may contribute to this difference. However, we disagree that item difficulty does not play a part. The figure shows a Rasch analysis which ranks items along a continuum according to item difficulty from very difficult to not difficult.¹ From this analysis it becomes clear that the MHAQ includes the less (or least) difficult item in five of the eight scales (rise, eat, walk, grip, reach). Only for one scale (dress) did the MHAQ use the more difficult item. In two other scales with three questions (hygiene, activities), the MHAQ used the item of intermediate difficulty. Because the MHAQ includes less difficult items, its score is systematically lower than the HAQ score.

We agree that, at present, there is no reason to favour the use of the MHAQ. Nevertheless, we believe that it would be worthwhile to develop a short version to encompass the full spectrum of item difficulty. This new version should, ideally, have interval-like characteristics.² A shorter version may increase patient compliance, and interval characteristics^{2,3} would require fewer



HAQ and MHAQ item difficulty. Brackets indicate HAQ scales. Horizontal markers indicate HAQ items and circles indicate MHAQ items.