Development of a functional scoring system for rheumatoid arthritis patients with cervical myelopathy

Adrian T H Casey, J Martin Bland, H Alan Crockard

Abstract

Objective—To be able to measure disability objectively in rheumatoid arthritis complicated by cervical myelopathy.

Methods—The responses to the Stanford health assessment questionnaire disability index were recorded from 250 consecutive patients (group 1) referred to our unit for spinal surgery. Using principal components analysis, the questionnaire was reduced from 20 questions to 10 questions. In the second part of the study, the responses of the questionnaire for those patients undergoing surgery from the original group of 250 patients were analysed with respect to outcome.

Results—The reduction in the number of questions resulted in no significant loss of information, reliability (internal consistency Cronbach’s α = 0.968) or sensitivity. The new scale, the myelopathy disability index, measures only one dimension (Eigen value 6.97) and may be more finely tuned to the measurement of disability in these myelopathic patients. When administered to the 194 patients undergoing cervical spine (group 2) surgery the myelopathy disability index was an accurate predictor of neurological and functional outcome, as well as survival following surgery (P < 0.0001).

Conclusions—The myelopathy disability index provides a much needed objective and reliable means of assessing disability in patients with rheumatoid involvement of the cervical spine and also in predicting outcome following surgical intervention. It also provides information for both the patient and surgeon alike, on what to realistically expect from surgery. Its adoption should facilitate comparisons between different forms of surgical intervention.

In determining the severity of any disease process, the effects of surgical intervention, or the factors that influence prognosis and survival, it is essential to have objective and reproducible means of measuring the patient’s disability before and after surgery. Pain can be reliably assessed by means of a visual analogue scale.1 Various clinical grading systems have been used to describe the neurological (Ranawat classes I-IIIIB)2 and functional (Steinbrocker’s grades I-IV)3 status of patients with rheumatoid arthritis with varying degrees of acceptance by orthopaedic and rheumatological circles (table 1). The Ranawat classification, while simple in design, fails to differentiate reliably between classes II and IIIA, the area of most clinical interest to the operating surgeon. What are the objective signs of myelopathy? How reliable is a neurological examination?4,5 Can one make valid judgements about power or tone in a patient with a painful arthritis where even passive movement is painful? Is it always possible to attribute the relative impact of a peripheral neuropathy or high cervical myelopathy on the symptoms of numb clumsy hands so common in patients with rheumatoid atlantoaxial subluxation? In these circumstances a functional scale might be more useful. This is more relevant to the patient’s needs, as what is interest is it to the patient that his reflexes are no longer brisk and his plantar response is downgoing following surgery if his level of disability remains unchanged? The American College of Rheumatology (ACR) has adopted the Steinbrocker’s classification but this is so crude that it has largely been dismissed by present day rheumatologists,4 although recent modifications have arguably improved its ability to measure patient satisfaction.

Interest in the measurement of physical function of chronically disabled patients with arthritis has increased significantly over the last 10 years. There is unfortunately no gold standard. Of the instruments developed, the physical disability dimension of the Stanford health assessment questionnaire (HAQ)7 has been shown to have the necessary qualities to evaluate functional disability and has been validated in many settings, with cross cultural reliability being demonstrated in several studies.8,9,10 Its use has largely centred on quantifying short term drug effects,11 but it also has been used to identify risk factors for disease progression.12,13 More recently it has also been modified for use with anklylosing spondylitis, systemic lupus erythematosus, diabetes, HIV, and orthopaedic evaluation following hip arthroplasty.14,15 It has not previously been validated specifically for cervical myelopathy secondary to the rheumatoid process. The purpose of this study has been to explore its use in this specific context, but also to see if refinements can be made to its structure by reducing the number of questions. This might help to reduce the level of noise or interference from extraneous questions and also make it more user friendly. An information...
Table 1 Ranawat neurological disability classification and The American College of Rheumatology grading system (A.C.R.) for functional disability

<table>
<thead>
<tr>
<th>Ranawat's neurological classification</th>
<th>Steinbrocker's grading system for functional disability</th>
</tr>
</thead>
<tbody>
<tr>
<td>I - No neurological deficit (normal neurological condition)</td>
<td>I - Complete ability to carry out all the usual duties without handicaps</td>
</tr>
<tr>
<td>II - Subjective weakness with hyperflexia and dysaesthesia</td>
<td>II - Adequate for normal activities despite handicap of discomfort or limited motion of one of the joints</td>
</tr>
<tr>
<td>III A - Objective weakness and long tract signs but able to walk</td>
<td>III - Limited to little or none of the duties of usual occupation or self-care</td>
</tr>
<tr>
<td>III B - Quadriparietic and non-ambulatory</td>
<td>IV - Incapacitated, largely or wholly bed-ridden or confined to a wheelchair with little or no self-care</td>
</tr>
</tbody>
</table>

Methods

PART I: RELIABILITY

In the first part of this study the preoperative responses of 250 patients (group 1) with rheumatoid arthritis to the Stanford health assessment questionnaire (self administered) were digitised and summed to yield an HAQ disability score. The properties of the questionnaire were first assessed using an inter-item correlation matrix, calculating its reliability index and the number of dimensions it measures by principal components analysis on the preoperative responses to the questionnaire in this group of patients. These 250 patients were recruited over a 10 year period (1983-1993), and form part of a consecutive series of rheumatoid arthritis patients referred for surgical evaluation because of significant radiological cervical spine abnormalities, occipitocervical pain, or the development of myelopathy. Their characteristics are summarised in table 2. The questionnaire was systematically reduced by removing those questions with a low variance, which therefore contained less discriminatory information than the others.

PART II: VALIDITY AND RESPONSIVENESS

In the second part of this study, those patients from the original group (table 2) who underwent surgery (group 2) during the same period were used to validate the new myelopathy disability index by analysing how well the new scale performed in predicting functional and neurological outcome at one year, in comparison to the previously used scales.

Kendall's rank correlation was used because the Ranawat and Steinbrocker classification are four point scales. Thus they cannot be from normal distributions and the product-moment correlation (Pearson's) is not appropriate. There are also likely to be many ties, so the Kendall coefficient is preferred to the Spearman's coefficient. Kaplan-Meier survival curves were also constructed to determine the new scale's relevance to survival following surgery. Construct validity was assessed by an expert panel. Biological validity (criterion related construct validity) was determined by correlations with known important radiological variables.

Reduction in the number of questions by removing those questions with low variance will obviously help in discriminating between individuals. However, it could be argued that this method might result in valuable loss of information. It is possible that some of the measures with the least variance preoperatively are the most responsive to surgery. For example, a low variance may be due to the fact that most patients are unable to perform a function on the original Stanford health assessment questionnaire. However, improvement in that item might be a responsive consequence of surgery. To determine whether question reduction resulted in loss of sensitivity or responsiveness, sensitivity statistics (the standardised response mean, SRM) and the effect size, ES were calculated for the original Stanford health assessment questionnaire and our modified instrument. These statistics calculate the magnitude of change in comparison with the standard deviation of change (for the SRM) or the standard deviation at the baseline (for the ES).

Results

The statistical properties of the questionnaire were explored by examining the level of inter-item correlation amongst the responses with all showing similar intercorrelations (average inter-item covariance 0.61). On the preoperative study group (group 1), consisting of 250 patients, the scale reliability coefficient (internal consistency) was 0.968.

Principal components analysis on the original 20 questions revealed that there was predominantly one component or dimension being measured (Eigen value 12.53) which explained 62.7% of the interperson variability. This dimension is essentially the sum of all the variables and our figure is similar to that recorded by Fries in the original assessment of his questionnaire, 65%. The second principal component in our patient population was smaller with an Eigen value of 1.24, accounting for 6% of interperson variability in comparison to Fries original study of 10%.

Having defined the properties of the original full version of the HAQ (20 questions) the next step involved reducing the number of variables. This was achieved by removing the variables with the lowest variance as they contain less information. Arbitrarily those variables with a standard deviation less than 1.0 were removed. This left 12 variables, with an average inter-item covariance of 0.74 and a scale reliability coefficient (Cronbach's α-internal
Functional scoring system for cervical myelopathy in RA

Table 3

PLEASE TICK THE ONE IN RESPONSE WHICH BEST DESCRIBES YOUR USUAL ABILITIES OVER THE PAST WEEK

<table>
<thead>
<tr>
<th></th>
<th>Without ANY difficulty</th>
<th>With SOME difficulty</th>
<th>With MUCH difficulty</th>
<th>Unable to do</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Rising are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stand up from an armless straight chair?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Get in and out of bed?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eating are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cut your meat!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lift a full cup or glass to your mouth?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walking are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walk outdoors on flat ground?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Climb up five steps?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hygiene are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wash and dry your entire body?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Get on and off the toilet?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grip are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open jars which have previously been opened?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activities are you able to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Get in and out of car?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

Note: If aids or assistance from another is required to perform any of the above tasks please score the activity as "with much difficulty". Total score = A+B+C+D (range 0-30). The final score is expressed as a percentage.

consistency) of 0.959. Two further items were removed for poor content validity by our expert panel. The question "Are you able to reach and get down a 5 lb object from above your head?" was not felt to be relevant, while the question "Are you able to open car doors?" depended too much on the type and date of car manufacture and introduced possible sources of inconsistency.

The remaining 10 questions of the HAQ are now referred to as the myelopathy disability index (table 3). This version had an average inter-item covariance of 0.74 and a scale reliability coefficient (internal consistency) of 0.95. The correlation between the original twenty question HAQ and the myelopathy disability index was $r = 0.98$, suggesting good preservation of information and providing objective evidence for the validity of this new scale.

Principal components analysis revealed that 70% of all the interpatient variability was explained by the first principal component (Eigen value 6.97) and the second component in the myelopathy disability index had an Eigen value of less than 1 (0.65). This represents an improvement on the original disability index.

From this sum (first principal component), we infer that the myelopathy disability index (an equal weight sum) is appropriate for measuring overall disability and the questionnaire is well focused for the measurement of this disability.

The distribution of the new scale (skew $-0.28$, kurtosis 1.96) more closely approximates to a normal distribution (fig 1) than the original health assessment questionnaire (skew $-0.55$, kurtosis 2.24).

CRITERION VALIDITY—CONCURRENT AND PREDICTIVE

The myelopathy disability index was then studied by seeing how it related to the preoperative and postoperative Steinbrocker's grades' and the Ranawat classes' (table 1) at 1 year, administered concurrently with the new disability index to the patients who ultimately underwent surgery, that is, group 2.

There were substantial rank correlations between the myelopathy disability index and the other pretreatment assessments. As might be expected, the correlations with postsurgical outcome measures (Ranawat neurological class, Steinbrocker's functional grades, and the Stanford HAQ) were smaller but still about 0.5 (0.48 to 0.53) with $P < 0.0001$ (table 4). Thus the myelopathy disability index is a predictor of outcome. It predicts the Ranawat neurological class and the Steinbrocker grades slightly better than the HAQ, though the difference is small. Moreover it predicts the post-treatment

Table 4 Correlation matrix (Kendall's rank correlation coefficients) documenting the relation between the pre- and postoperative status of the patients according to the commonly used grading systems. All correlation are highly significant at $P < 0.0001$

<table>
<thead>
<tr>
<th>Grading system</th>
<th>Myelopathy disability index</th>
<th>Health activity questionnaire (HAQ)</th>
<th>Ranawat classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preoperative</td>
<td>0.92</td>
<td>0.62</td>
<td>0.63</td>
</tr>
<tr>
<td>Ranawat classification</td>
<td>0.63</td>
<td>0.62</td>
<td>NA</td>
</tr>
<tr>
<td>Steinbrocker's (ACR) grades</td>
<td>0.70</td>
<td>0.69</td>
<td>0.71</td>
</tr>
<tr>
<td>Postoperative</td>
<td>0.53</td>
<td>0.54</td>
<td>0.47</td>
</tr>
<tr>
<td>Ranawat classification</td>
<td>0.48</td>
<td>0.47</td>
<td>0.51</td>
</tr>
<tr>
<td>Steinbrocker's (ACR) grades</td>
<td>0.50</td>
<td>0.48</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Figure 1 Histogram illustrating the distribution of the myelopathy disability index score, expressed as a percentage. The normal distribution curve is superimposed.
scores in a similar way to the pretreatment Ranawat class (fig 2).

RESPONSIVENESS/SENSITIVITY STATISTICS
The myelopathy disability index is sensitive or responsive to change. The standardised response mean (SRM) was 0.36, with a similar value for the effect size (ES), 0.39. These values have been classified as moderately sensitive. This therefore makes it a suitable instrument to measure the effects of surgical intervention. There is a floor effect. Using either scale (MDI or HAQ)—27/194 of our patient population undergoing surgery have maximum disability according to the MDI; this figure is unchanged when using the parent instrument (Stanford HAQ). The sensitivity statistics for the Stanford health questionnaire for our patient population are very similar to those of our abbreviated questionnaire, myelopathy disability index—SRM −0.40; ERM −0.31. Questionnaire reduction has not therefore resulted in any loss of sensitivity or responsiveness. The corresponding values for the Steinbrocker grades (SRM −0.32; ERM −0.35) and Ranawat classes (SRM −0.43; ERM −0.42) were essentially similar. These gradings, however, are only 4 point scales and while they may have similar sensitivity statistics to the summed HAQ and MDI it should not be forgotten that the changes in individual question responses before and after surgery are often of interest to the clinician.

SURVIVAL ANALYSIS
Kaplan–Meier survival curves were created to examine survival following surgery (fig 3). A Cox regression model was then constructed to examine the predictive power of the various grading systems in predicting survival following surgery. Both the HAQ (P < 0.0002), and the myelopathy disability index (P < 0.0001), are significant predictors of survival, the latter predicting slightly better, as shown by a higher ratio of coefficient to standard error and hence lower P value and the higher pseudo R². The Steinbrocker and the Ranawat grades are also good predictors of survival (P > 0.0001) as shown in fig 3.

CONSTRUCT VALIDITY—RADIOLOGICAL CORRELATES
Finally, there are objective radiological measurements that correlate with the development of cervical myelopathy described in published reports. The applicability of this new functional disability index was examined by calculating its correlation with important radiological indices including spinal canal diameter (posterior atlantodens interval, r = −0.25, P = 0.0009), spinal cord diameter (r = 0.28, P = 0.0009), spinal cord area (r = −0.34, P < 0.0001) at the site of compression, and...
the degree of vertical translocation of the odontoid peg through the foramen magnum \( r = -0.29, \ P = 0.0002 \). All correlated significantly, and in the appropriate direction, with the new myelopathy index, albeit to varying degrees.

Discussion

The advantages of modifying an existing questionnaire is that the questions have already gone through a process of evolution and rigorous assessment. Indeed it is on this very basis that Fries and his colleagues set about creating the Stanford health assessment questionnaire with potential component questions being selected from a variety of sources including the Bartel index and the index of independence in activities of daily living. Areas of redundancy, imprecision, and ambiguity were identified and the questionnaire appropriately modified. By adopting their questions we have obviated the need for formal retesting of these questions which have already undergone tests of replicability or reliability, objectivity, and validity. The use of a self administered questionnaire has advantages over an interview as a means of assessing outcome. It is less expensive to administer and does not depend on the skill of a particular interviewer or require interobserver validation.

The limitations of a neurological examination in the determination of mild muscle weakness have been highlighted in a non-arthritis patients with a kappa coefficient of only 0.22 when the examiner was unaware of the clinical history. Indeed the major difficulties presented by patients with rheumatoid arthritis with their painful deforming arthritis, which so limits a conventional neurological examination was one of the major forces in the development of our myelopathy disability index. We would be naive to believe that our statistical manoeuvres could completely separate functional disability as a direct consequence of cervical cord compression from the general effects of the rheumatoid disease process, but we feel that our disease specific instrument will discriminate between myelopathic patients in terms of their physical disability.

One of the criticisms levelled against the original HAQ was an undue weighting for upper limb function which our statistical manoeuvres have corrected. Our new myelopathy disability index incorporates four questions dealing with upper limb function, four dealing with lower limb function, and the remaining two relate to a combination of both. This reduced questionnaire has improved qualities, based purely on statistical considerations, as witnessed by principal component analysis which reveals only one dimension to our new disability index and a distribution that more closely approaches normal than the original HAQ disability index, in combination with a high reliability coefficient. A previous attempt at HAQ reduction did not succeed because the addition of transition questions to improve the instrument's ability to detect change unfortunately rendered the modification too complex.

On a clinical basis we have established that the myelopathy disability index performs at least as well, if not better than, the existing disease severity and outcome measures currently used in clinical practice. Its has major practical advantages over the Ranawat neurological classification and the Steinbrocker grading system, neither of which have been validated either statistically or clinically. These advantages include a better differentiation of disease state and sensitivity to change with the expansion of grades from 4 points on the Ranawat and Steinbrocker systems to 30 points (expressed as a percentage) on the myelopathy disability index.

We have also demonstrated that this scale is a powerful predictor of outcome following surgery, both in terms of neurological and functional outcome but perhaps more importantly of survival following surgery. It has clinical relevance in that it has significant correlations with those radiological variables of interest (posterior atlantodens interval, cord diameter, cord area, and the degree of vertical translocation). Indeed for a recent review, see references 14-16.

An area of future research would seek to identify those patients, using this new myelopathy score, with neurological deterioration which hitherto has been masked by their painful deforming arthritis. A deterioration in function, independent of a flare up in their rheumatoid process, would be an indication for the clinician to reassess the patient both clinically and radiologically (x rays, CT scan, and MRI) with a particularly critical eye. We have found this, in conjunction with a pain visual analogue scale for occipitocervical pain, to be a useful adjunct to our standard clinical evaluation in detecting late failures in our surgical series. This of course measures a different dimension and could not be incorporated directly into our disability index.

Surgery to the rheumatoid cervical spine is not a minority interest, it is a major health problem, with a recent cross sectional population study from Finland suggesting that 10% of patients with rheumatoid arthritis fulfil the accepted standard criteria for operative intervention. In the USA, with an overall prevalence of 0.9%, there are an estimated 2.2 million patients with definite rheumatoid arthritis, extrapolation of the provocative and controversial Finnish data would suggest a pool of 220 000 potential surgical candidates.

This new myelopathy disability index scale would have widespread applicability and its ease of administration would not confine it to academic centres alone. The lack of a common language of reliable and well validated instruments has produced a situation in which it has become difficult to assess the results of treatment objectively, or document the effectiveness or indeed ineffectiveness, of surgical intervention. The whole area of outcome research in cervical spine surgery has been found to be wanting with little emphasis on patient-perceived outcomes. This new disability index will hopefully remedy this shortcoming for rheumatoid cervical myelopathy, but future uses may also include extrapolating the use of this myelopathy score to other causes of cervical
myelopathy such as cervical spondylosis, spinal cord trauma, or tumour surgery. Its adoption in these areas will no doubt complement the existing methods of assessment.33-35

CONCLUSION
In conclusion, we have refined the Stanford HAQ for use in rheumatoid cervical myelopathy. We believe this to be a superior instrument to the original HAQ for specific purposes of measuring disability in rheumatoid patient with cervical myelopathy. Combined with a pain visual analogue scale it provides a comprehensive assessment of the patient with rheumatoid involvement of the cervical spine. The myelopathy disability index has been shown in this study to convey useful prognostic information and has the potential to form the basis of future outcome studies comparing the results of surgery between different surgical groups and management options.

ATHC was a recipient of Grant Support from the Welton Foundation.

Development of a functional scoring system for rheumatoid arthritis patients with cervical myelopathy.
A T Casey, J M Bland and H A Crockard

*Ann Rheum Dis* 1996 55: 901-906
doi: 10.1136/ard.55.12.901

Updated information and services can be found at:
http://ard.bmj.com/content/55/12/901

*These include:*

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/