Cross-cultural adaptation, reliability and validity of the German Shoulder Pain and Disability Index (SPADI)

F. Angst, J. Goldhahn¹, G. Pap, A. F. Mannion¹, K. E. Roach², D. Siebertz, S. Drerup¹, H. K. Schwyzter and B. R. Simmen

Objective. To cross-culturally adapt the Shoulder Pain and Disability Index (SPADI) from English into German, and to test the reliability and validity of the German version.

Methods. Cross-cultural adaptation of the SPADI was performed according to international guidelines. One hundred and eighteen patients who had undergone shoulder arthroplasty, on average 4 yr previously, completed a questionnaire booklet containing the German SPADI, the Short Form 36 (SF-36), the Disability of the Arm, Shoulder and Hand (DASH) questionnaire, and the American Shoulder and Elbow Surgeons (ASES) questionnaire for the shoulder to assess SPADI's construct validity. One week later, they completed the SPADI again to assess test–retest reliability.

Results. The six-step cross-cultural adaptation procedure revealed no major problems with the content or language. The intraclass correlation coefficients for the individual items of the SPADI were between 0.68 and 0.89, and that for the SPADI total score was 0.94. The SPADI total score showed a correlation of 0.61–0.69 with the SF-36 physical scales, of 0.88 with the DASH and of 0.92 with the ASES.

Conclusions. The German SPADI is a practicable, reliable and valid instrument, and can be recommended for the self-assessment of shoulder pain and function.

Key words: Shoulder, Questionnaire, Reliability, Validity, SPADI.

The shoulder is one of the most common sites of musculoskeletal pain, second only to back pain, with 1-month to 1-yr period prevalences in the general population of 6–37%. Shoulder problems are hence associated with considerable personal and economic costs in relation to the individual sufferer and the healthcare system alike [1]. It is becoming apparent that the patient’s subjective perception of their disease status is decisive for both the diagnostic work-up and the subsequent therapeutic management [2]. Consequently, compared with the evaluation of clinical and biomechanical factors, self-assessment is gaining increasing importance in determining the functional and psychosocial (participation) consequences of disease (World Health Organization International Classification of Functioning, Disability, and Health: ICF levels 2 and 3) [4].

The shoulder consists of five functional joints, which form a functional entity to perform complex movements in almost every activity of daily living [1]. A joint-specific assessment of shoulder disorders should cover a range of functions and movements of this functional entity that is as wide as possible but exclude functions of the elbow and wrist as far as possible. The WHO’s ICF demands thoroughness and most commonly used self-assessment instruments for the shoulder [8, 9]. The DASH was originally developed for the assessment of all three joints of the upper extremity and hence is not shoulder-specific. The ASES (patient self-assessment part), a questionnaire that resulted from a Delphi process carried out amongst shoulder surgeons, is longer than the SPADI (18 items vs 13 items) and has not been tested extensively. A German version of the DASH already exists [10] and that of the ASES will soon be published [J. Goldhahn, F. Angst, G. Pap, A. F. Mannion, S. Drerup, H. K. Schwyzter and B. R. Simmen; [11]. Previous studies have shown that it is feasible to implement the English original version of the SPADI in clinical practice and that it is a valid and highly responsive instrument for assessing shoulder pain and function [11–13]. The aim of this study was to cross-culturally adapt the SPADI for use with German-speaking patients and to determine its clinimetric properties, in particular, the reliability and validity.

Materials and methods

Translation and cross-cultural adaptation

The SPADI comprises a series of 13 items (five for pain, eight for function), each scored with a visual analogue scale (VAS) from 0 (no pain/no difficulty, i.e. best) to 11 (worst pain imaginable/some
difficult required help, i.e. worst) [7]. We followed the description of the original SPADI and adopted this likely unusual scaling ‘A numeric score was calculated for each item by arbitrarily dividing the horizontal line into 12 segments of equal length. A number ranging from 0 to 11 was attached to this segment to produce a score for each item’ [7]. For the score transformation, there is no problem whether the item maximum is scored 10 or 11: the divisor is 11 instead of 10.

Cross-cultural adaptation of the SPADI was performed following the guidelines of the American Association of Orthopedic Surgeons (AAOS) Outcomes Committee [14]. This process consists of six steps, each of which is documented with a written report: stage I, forward translation (English to German) by an informed translator (i.e. a health professional, T1) and uninformated translator (T2); stage II, synthesis of T1 and T2, resolving any discrepancies, leading to version T-12; stage III, back-translation (German to English) of the version T-12 by two native English-speaking back-translators (BT1 and BT2), naive to the purpose of the instrument; stage IV, expert committee review consisting of all four translators, one methodologist and one language professional to reach consensus on discrepancies or ambiguities, and to establish a pre-final version (German); stage V: pretesting of the target language version to examine the layout, wording, ease of understanding and ease of completion of the questionnaire; stage VI, submission of the documentation of stages I–V to the developers of the original questionnaire or the AAOS to ensure that the process has been carried out correctly and that a reasonable translation has been achieved.

Subjects

The reliability and validity of the SPADI were examined by a catamnesis of a group of patients who had undergone primary shoulder arthroplasty (mostly due to osteoarthritis or rheumatoid arthritis) and who had been invited to attend for a follow-up examination, on average 4.0 yr after their operation. The first questionnaire was completed during the visit to the hospital and a second questionnaire was completed by the patient at home, 1 week later, and returned by mail. The study protocol was approved by the local ethics committee and written informed consent was obtained from all participants.

Reliability

The intraclass correlation coefficient (ICC) was used to quantify test–retest reliability or stability over time, i.e. the extent to which the same test results are obtained for repeated assessments when no real change is expected in the intervening period. The ICC was calculated for the agreement between the two (test and retest) responses for each item in the questionnaire, for the SPADI subscales (pain and disability) and for the total SPADI score. The ICC can range from 0.00 (no agreement) to 1.00 (perfect agreement). Cronbach’s alpha (CA) is computed as the average correlation (0.00, no correlation; 1.00, perfect correlation) between the items of a scale to quantify the internal consistency of the item(s) within the scale. Detailed explanations of the ICC and CA can be found in [15, 17].

Several indicators of the precision (or the error) of measurement, as additional indices of reliability, can be found in the literature. One such measurement is the ‘standard error of measurement’ (SEM; not to be confused with the standard error of the mean in statistics) and another is the ‘typical error’ (TE) [17]. The SEM is equal to the standard deviation of the group’s scores divided by the square root of (1 – CA) in the one-point measurement case and of (1 – ICC) in the test–retest data case, as used for our data [15, 18]. The typical error (TE) is equal to the standard deviation of the differences between the test and retest measurements for each patient (sΔ) divided by √2; it is also given by the square root of the within-person residual mean square from the analysis of variance (ANOVA) of the test and retest results [17]. The 95% confidence interval (CI) for these measurements can be used to quantify the precision of measurement as the mean of the difference (mΔ); mΔ ± tn–1,0.025 × SEM [18] or mΔ ± tn–1,0.025 × sΔ [17], where tn–1,0.025 is the t-value of the Student’s t-distribution and sΔ = TE × √2. Further explanations of the two concepts can be found in the Results.

Validity

The validity of the SPADI was examined by determining how well the scores correlated with those of the Short Form 36 (SF-36), the DASH and the ASES. The results were compared with those previously reported in the literature, as revealed by a PubMed/Medline review of the primary and linked secondary literature for the three most commonly used shoulder instruments, the DASH, the ASES and the SPADI, and for the generic SF-36 [19, 20]. The SF-36 assesses physical, mental and psychosocial health and ability [21]. An overview of the most important shoulder assessment instruments can be found in [8, 9]. The version of the German SPADI used in our previous validity study was identical to that presented in this study [21].

The following concepts have been well defined and explained [22, 23]. Face or content validity is concerned with how well the items or questions seem to make sense; i.e., in the case of the SPADI, how relevant they seem to be to the assessment of shoulder pain and function. Criterion validity describes the extent of agreement between the measure of interest (i.e. the SPADI) and the gold standard (the criterion), i.e. real shoulder pain and function. In the absence of a gold standard, the correlation between the scores on the new instrument and those on a measure that is similarly constructed and has already been validated (e.g. the DASH and patient ASES) can be examined, to determine the extent to which the new instrument corresponds to the theoretical concept of shoulder pain and function (construct validity). The criterion and the construct validity are usually quantified by non-parametric Spearman rank correlation coefficients, as the latter do not require that the data be distributed (statistically) in any specific manner. High correlations are expected between instruments with similar constructs (e.g. the SPADI and the ASES), proving convergent validity and low correlations between instruments with different constructs (e.g. the SPADI and the SF-36) proving discriminant validity [16].

Data were collected using standardized evaluation forms that were then scanned into a database. Prior to analysis, the five pain items and the eight function items (scaled from 0 = best to 11 = worst) of the SPADI were transformed to a scale ranging from 0 (worst) to 100 (best) for ease of comparison with previous studies in which scores were expressed on a 0–100 scale [21]. For example, the pain score is given by (11 – mean)/11 (worst) of the SPADI were transformed to a scale ranging from 0 (worst) to 100 (best) for ease of comparison with previous studies in which scores were expressed on a 0–100 scale [21]. For example, the pain score is given by (11 – mean)/11×100, where mean is the mean of the five pain items. All analyses were carried out using the statistical software package SPSS 11.0 for Windows® (SPSS, Chicago, IL, USA).

Results

Cross-cultural adaptation

The original English version is shown in Table 1 and the final German version is available as supplementary data at Rheumatology Online. Forward and back-translation of the SPADI revealed no major problems or language difficulties. Most of the discrepancies in the forward and back translations (stages I–III) concerned synonyms for particular expressions. For example, ‘how severe is your pain’ (items 1–5): severe → stark → strong, ‘at its worst’ (item 1); worst → ‘am schlimmsten’, ‘removing something from your back pocket’ (item 13): removing → ‘aus…nehmen’ → ‘take…out’, etc.
Table 1. Descriptive and reliability data for the German SPADI in patients >1 yr after hemi-arthroplasty or total shoulder arthroplasty (n = 125 joints)

<table>
<thead>
<tr>
<th>Item</th>
<th>Wording of the original SPADI [7]</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>s.d.</th>
<th>ICC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>How severe is your pain:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>02</td>
<td>At its worst</td>
<td>0 (22%)</td>
<td>11 (2%)</td>
<td>3.9</td>
<td>3.2</td>
<td>0.76</td>
<td>0.68–0.83</td>
</tr>
<tr>
<td>03</td>
<td>When lying on the involved side</td>
<td>0 (39%)</td>
<td>11 (2%)</td>
<td>2.9</td>
<td>3.4</td>
<td>0.86</td>
<td>0.81–0.90</td>
</tr>
<tr>
<td>04</td>
<td>Reaching for something on a high shelf</td>
<td>0 (30%)</td>
<td>11 (6%)</td>
<td>3.4</td>
<td>3.6</td>
<td>0.75</td>
<td>0.66–0.82</td>
</tr>
<tr>
<td>05</td>
<td>Touching the back of your neck</td>
<td>0 (50%)</td>
<td>11 (6%)</td>
<td>2.4</td>
<td>3.5</td>
<td>0.87</td>
<td>0.82–0.91</td>
</tr>
<tr>
<td>06</td>
<td>Picking up the involved arm</td>
<td>0 (34%)</td>
<td>11 (5%)</td>
<td>2.7</td>
<td>3.3</td>
<td>0.81</td>
<td>0.74–0.86</td>
</tr>
<tr>
<td>07</td>
<td>How much difficulty do you have:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>Washing your hair</td>
<td>0 (35%)</td>
<td>11 (10%)</td>
<td>3.3</td>
<td>3.9</td>
<td>0.89</td>
<td>0.85–0.92</td>
</tr>
<tr>
<td>09</td>
<td>Washing your back</td>
<td>0 (19%)</td>
<td>11 (19%)</td>
<td>5.3</td>
<td>4.3</td>
<td>0.83</td>
<td>0.76–0.88</td>
</tr>
<tr>
<td>10</td>
<td>Putting a shirt that buttons down the front</td>
<td>0 (54%)</td>
<td>11 (1%)</td>
<td>1.8</td>
<td>2.7</td>
<td>0.68</td>
<td>0.57–0.76</td>
</tr>
<tr>
<td>11</td>
<td>Putting on your pants</td>
<td>0 (61%)</td>
<td>11 (2%)</td>
<td>1.5</td>
<td>2.6</td>
<td>0.77</td>
<td>0.69–0.84</td>
</tr>
<tr>
<td>12</td>
<td>Placing an object on a high shelf</td>
<td>0 (18%)</td>
<td>11 (17%)</td>
<td>4.9</td>
<td>4.0</td>
<td>0.82</td>
<td>0.75–0.88</td>
</tr>
<tr>
<td>13</td>
<td>Carrying a heavy object of 10 pounds</td>
<td>0 (26%)</td>
<td>11 (16%)</td>
<td>4.8</td>
<td>4.2</td>
<td>0.89</td>
<td>0.85–0.92</td>
</tr>
<tr>
<td>14</td>
<td>Removing something from your back pocket</td>
<td>0 (41%)</td>
<td>11 (4%)</td>
<td>2.6</td>
<td>3.3</td>
<td>0.73</td>
<td>0.63–0.80</td>
</tr>
<tr>
<td>15</td>
<td>Function</td>
<td>0 (1%)</td>
<td>100 (14%)</td>
<td>72.2</td>
<td>26.7</td>
<td>0.89</td>
<td>0.84–0.92</td>
</tr>
<tr>
<td>16</td>
<td>SPADI (total score)</td>
<td>3</td>
<td>100 (7%)</td>
<td>71.0</td>
<td>25.2</td>
<td>0.94</td>
<td>0.91–0.96</td>
</tr>
</tbody>
</table>

Descriptive data: minimum, maximum, mean and s.d. are those from the first administration of the questionnaire; the descriptive data for the retest results are not reported. The percentages in parentheses indicate the frequencies of floor and ceiling values (see text).

95% CI, 95% confidence interval for the ICCs.

Item 4. 'How severe is your pain when touching the back of your neck', needed consensus discussion (stage IV). In German, the word ‘Nacken’ is used for the back of the neck. However, this was then back-translated as simply ‘neck’ rather than ‘back of the neck’ (i.e. ‘how severe is your pain when touching your neck’). Strictly speaking, this translation is not incorrect, and probably arose as a result of the back-translators being unaware of the purpose of the questionnaire. ‘10 pounds’ (item 12), which is actually 4.54 kg, was expressed as ‘etwa 5 kg’ (approximately 5 kg) for ease of understanding. Finally, consensus had to be gained in the translation of ‘putting on a shirt that buttons down the front’ (item 9) by ‘wenn Sie ein Hemd oder eine Bluse anziehen, die vorn zugeknöpft wird’, since ‘that buttons down the front’ cannot easily be translated word by word, and back-translation resulted in ‘which has buttons up the front’.

The extreme (worst) anchor for pain resulted in ‘schlimmste Schmerzen’ (worst pain) instead of ‘worst pain imaginable’ because ‘schlimmste vorstellbare Schmerzen’ was considered a strange formulation and would have been difficult to understand. The extreme (worst) anchor for function (‘so difficult that required help’) was translated as ‘Tätigkeit nicht ausführbar’ (activity not feasible/unable to do), which was considered to be synonymous with the original response category but easier to judge (because availability and sources of help may differ between patients) and analogous to other instruments, such as the DASH. Pretesting of the pre-final German version (stage V) revealed no further difficulties with the questionnaire in nine patients randomly selected from the hospital waiting room, five health professionals and 42 participants in an earlier, related study [21]. The cross-cultural adaptation process and the final version were approved by K. E. Rouach, who developed the English SPADI in 1991 (stage VI) [7].

Subjects
A total of 176 German-speaking patients, who had undergone total shoulder or hemi-arthroplasty of the shoulder at least 1 yr previously (for rheumatoid arthritis or osteoarthritis), were invited to attend the hospital for a follow-up assessment, as part of a larger clinical study (to be published separately). Thirty-two patients (18.2%) could not be assessed: seven had died, 15 were too ill to attend the hospital or to fill in a questionnaire (cancer, dementia, etc.), five had moved to an unknown address and five refused to participate. Of the 144 patients examined, 23 (16.0%) returned incomplete data, i.e. filled in fewer than 2/3 of the items in the pain or function scales, which contravened the previously established missing data rule [21], and three were excluded between the test and retest examinations, due to injury or re-operation of the shoulder. On checking the missing data items, no special preference was detected which could hint, for example, that an item is specially difficult to understand.

This resulted in 118 patients with complete test and retest data (resulting in 125 questionnaire sets because seven patients had been operated bilaterally), examined 1.2–7.9 yr (mean 4.0 yr, s.d. 1.2 yr) after their operation. Their ages ranged from 32.7 to 88.9 yr (mean 69.2 yr, s.d. 11.6 yr) and 77 (65.3%) were female. Twenty out of 125 (16.0%) of all operations were hemi-arthroplasties and 105/125 (84.0%) total arthroplasties.

Reliability
Table 1 shows the descriptive data for the individual items, the subscales and the total SPADI scale, along with the test–retest ICCs. The items showed relatively high ceiling effects (i.e. item score = 0) up to 61% for item 10 (difficulty in putting on your pants)—whereas the floor effects (i.e. item score = 1) were moderate to low, with a maximum of 19% for item 7 (difficulty in washing your back). Consequently, none of the items or scores was normally distributed (Komolgorov–Smirnov test). The pain, function and total scores were around 70 points (100 = best) and showed very low floor and low ceiling effects.

Of all possible items, 47.7% (125 questionnaires × 13 items = 1625 values) had identical scores at the test and retest (integer item numbers/values from 0 to 11); the lowest agreement was found for item 1 (how severe is your pain at its worst? 37.6% identical values), and the highest for item 10 (difficulty in putting on your pants; 68.8% identical values). The item ICCs ranged from 0.68 to 0.89; the least reliable item was item 9 (difficulty putting on a shirt that buttons down the front) and the most reliable were items 6 (difficulty washing your back) and item 12 (difficulty in carrying a heavy object of 10 pounds).

Of the 13 items, eight showed ICCs greater than 0.80. The ICCs of the scales were 0.89 for pain, 0.93 for function and 0.94 for the SPADI total score, i.e. the higher the number of items in the scale,
the higher the ICC. Internal consistency as measured by the CA was 0.92 for pain, 0.93 for function and 0.95 for the SPADI total score.

Table 2 shows the results describing the precision of measurement. The values for the typical error (TE) indicate that when a given patient is tested and retested, a change of $t_{124,0.025} \times s_{\Delta} = 1.979 \times 9.11 \times 18$ points represents the minimum difference in the total SPADI score required to state (with 95% confidence) that a real change is responsible for the difference and not just measurement error, i.e. 18 is the minimum detectable change score (MDC95%) [17]. In contrast, the concept of Roddey (and Portney) interprets the interval of −15 to +10 as the range in which the examiner can be 95% confident that the individual’s true total score is located [15, 18].

Because of the relatively large ceiling effects, post hoc analysis was performed for all single items and scores that did not represent the minimum and maximum score possible (data not shown in detail). The items revealed ICCs from 0.60 (item 13) to 0.79 (item 6), with the exception of item 9, which had an ICC of 0.44. The ICC was 0.87 for the pain score, 0.91 for the function score and 0.93 for the total score.

### Validity

Face validity was established in the original English version of the SPADI and was also considered to be adequate for the German SPADI, following discussions within the expert committee (consensus committee at stage IV), i.e. the content of the translated items made sense in the assessment of shoulder pain and function. The Spearman rank correlation coefficients describing the extent of the correlation between the SPADI scores and those of the comparison questionnaires (i.e. to examine SPADI’s construct validity) are shown in Table 3, along with similar data reported in the literature for the original English SPADI [12, 13, 21]. The correlations between the SPADI and the SF-36 physical dimensions, and those between the SPADI and the SF-36 overall score were between 0.61 and 0.69; the correlations between the SPADI and the upper extremity specific measures were somewhat higher: 0.76–0.92 (all $P < 0.001$; Table 3).

### Discussion

The German SPADI proved to be a reliable and valid measure of shoulder pain and function that was feasible to implement in clinical practice. The questions are short, easily understandable and the whole form requires approximately 2 min to complete [21]. The reliability and internal consistency were excellent for both the subscores and the total score. The reliability was also good for the individual items, with the exception of item 9 (difficulty in putting on a shirt that buttons down the front), for which it was still moderate. Compared with the original English SPADI, the German SPADI showed similar or stronger correlations with the SF-36 physical dimensions, the DASH and the (patient) ASES.

The reliability of the German SPADI was higher (SPADI total score, ICC = 0.94) than that of the English version (ICC = 0.66 [7] and 0.91 [12]). The same is true for the subscores: German SPADI, ICC = 0.89 vs original SPADI, ICC = 0.63 for pain and ICC = 0.93 vs 0.64 for function [7]. According to Bot et al. [9], an ICC ≥0.90 allows the reliable assessment of individual patients. This figure was reached for the total score and the function subscore, and was almost reached by the pain subscore. The internal consistency of the German SPADI was equal to or slightly higher than that of the original SPADI: CA = 0.92 for pain, 0.93 for function and 0.95 for the total German SPADI vs 0.86, 0.93 and 0.95 [7], and 0.89, 0.95 and 0.96 [18] in the original SPADI. CA values greater than 0.90 are interpreted to mean that some items are probably redundant, which may limit the content validity [15].
The standard errors of measurement (SEM) of the German SPADI (Table 2; e.g. the SPADI total score, 6.2) were somewhat higher than those reported by Roddey et al. (e.g. SPADI total score, 4.8) [18]. The typical error (TE) of the German SPADI (total score, TE = 6.4) was higher than that reported for other musculoskeletal symptom-specific questionnaires, e.g. for the Oswestry (Low Back) Disability Index (TE = 3.4), which is also scaled 0 (worst) to 100 (best) [24]. This means that the German SPADI showed lower precision of measurement than the original SPADI and than another musculoskeletal pain and disability instrument. Notably, the concepts of Hopkins and Roddey led to different 95% CIs, which means that there is no generally accepted single method.

The reliability was still acceptable (reducing by just 0.00-0.21 coefficient points), even when the scores that represented a minimum or maximum possible value for a given item/sum-score were excluded. A sample with high floor and/or high ceiling effects will tend to overestimate the reliability, since a test value of 0 (e.g. best health, ceiling) is more likely to be rated identically at the retest than is a score in the middle of the range (e.g. item score 3 = low pain). To our knowledge, this post hoc analysis is not usually performed in most studies examining the psychometric properties of questionnaires, but we recommend that it be done in order to assess the stability of the reliability data.

The face and construct validities of the German SPADI were good, and comparable to (or slightly better than) that of the original SPADI (Table 3) [11, 13, 21]. As expected, because of the differences in the constructs, the correlations between the scores for the condition-specific SPADI and those of the generic SF-36 were low to moderate (proving discriminant validity), whereas the correlations between the SPADI scores and those of the condition-specific instruments (DASH, ASES) were (very) high (proving convergent validity) [16]. The correlations reported in the present study and in the literature to date were relatively consistent across the different settings (Table 3), except those between the SPADI and the patient ASES (r = 0.77, 0.81, 0.92); this should be verified by further studies.

The study’s strengths include the standardized methods used for all procedures, especially for the cross-cultural adaptation, and the large sample size. Assessment of cross-cultural adaptation by reliability and validity does not depend on specific diagnoses of shoulder affections (and by that it does not depend on a specific patient selection) since all comparisons are made within the same patient(s) (within-patient data comparisons) and not between different patients. Relevant limitations in relation to the generalizability of the results include the fact that the study was performed only in the German-speaking part of Switzerland and only in shoulder arthroplasty patients. In Germany and Austria, the same (written) version of SPADI is used as in Switzerland, and language differences between the countries relate only to the spoken dialects. Hence, we expect the questionnaire to perform just as well in these German-speaking countries, although this would have to be verified in further studies. The high ceiling effects observed (2–19% for the items)—perhaps a consequence of examining postoperative patients, in whom symptoms would be expected to be lower—may have inflated the scores for the reliability and validity of the questionnaire: a patient who scores 100 (no pain/no difficulty) in the first assessment would be more likely to score 100 again at retest, which results in perfect reliability for that patient. However, this did not appear to have any notable influence on the results, as post hoc analysis detected no large bias, and other studies that could also be expected to have been affected by similarly high ceiling effects showed comparable results (e.g. [11]).

From a clinical perspective, ceiling effects reveal important limitations of an assessment. Two patients with ceiling scores cannot be (further) differentiated or individually characterized by the instrument. This is not a problem in pain if both do not have pain, but one patient could have better function (e.g. able to lift 10 kg) than the other (e.g. able to lift 20 kg). Additionally, a patient with a ceiling score at baseline cannot improve but only remain stable or worsen at the follow-up. This phenomenon is mainly responsible for the regression-to-the-mean effect of effects measured by closed scales.

Previous studies have shown that, in assessing clinical change in shoulder patients, the English SPADI is more responsive than the SF-36, the modified ASES for the shoulder and four other shoulder-specific instruments [12]. Our future prospective studies will examine the sensitivity to change (responsiveness) and the minimal clinically important (perceptible) difference for the German SPADI, in order to complete the assessment of its clinimetric properties.

<table>
<thead>
<tr>
<th>Key messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>- The German SPADI showed excellent reliability and validity, with fair measurement precision.</td>
</tr>
<tr>
<td>- The German SPADI can be recommended for use in clinical studies for the condition-specific assessment of shoulder pain and function.</td>
</tr>
</tbody>
</table>

**Acknowledgements**

Funding to pay the Open Access publication charges for this article was provided by...

We gratefully thank Astrid Junge, Geoffrey Klein and Vivien Robinson for their help with the cross-cultural adaptation, Gordana Balaban and Katrin Knecht for their help with the data collection, and all the patients who participated in the study. This study was supported by the Georg und Bertha Schwyz-Winker Foundation, Zurich, Switzerland.

There are no competing interests to declare.

**Reference**

Clinical Vignette

An unusual association with Raynaud’s phenomenon

A 36-yr-old lady with a year of typical Raynaud’s and polyarthralgia had a normal examination other than cold peripheries and blood pressure of 90/60 mmHg and no evidence of connective tissue disease. Her haemoglobin was 10.3 g/dl and sodium 127 mmol/l (135–145), but other tests, including thyroid stimulating hormone (TSH) and immunology, were normal. Six months later she was reviewed and remarkably improved. She reported that in the intervening time an endocrinologist had investigated episodes of collapse, fatigue and dizziness. She had been well until the delivery of her only child 3 yrs ago. There was significant postpartum bleeding. Following this she lost her pubic hair, libido and was unable to breast feed. She was being investigated for premature menopause. She had an abnormal short synacthen test, low free T3 and T4, gonadotrophin and growth hormone. An MRI showed atrophy of the pituitary gland suggesting a previous infarction (Fig. 1) Panhypopituitarism (Sheehan’s syndrome) was diagnosed. On starting hydrocortisone, thyroxine and the oral contraceptive her Raynaud’s settled. We have only found one other similar case in the literature [1].

There is no conflict of interest or source of funding.

M ARIA A L-DEIRI¹, J ONATHAN J. B ENN², C HRIS M. D EIGHTON¹

¹ Derb yshire Ro yal In firmary, D epartment of Rheum atology, Derb y, Derb yshire and ² B urton Hospitals NHS T rust, D epartment of Medicine, Burton-on-Trent, Staffs ordshire, UK.

Correspondence to: C. M. Deighton.
E-mail: chris.deighton@derbyhospitals.nhs.uk

doi:10.1093/rheumatology/kel304

Fig. 1. Sagittal MRI scan with contrast demonstrating the pituitary atrophy (arrow).