Computer-Administered Bath Ankylosing Spondylitis and Quebec Scale Outcome Questionnaires for Low Back Pain: Agreement with Traditional Paper Format

HELEN BENT, CHARLES R. RATZLAFF, EWAN C. GOLIGHER, JACEK A. KOPEC, and JEAN H. GILLIES

ABSTRACT. ***Objective***. To measure the agreement between computer and paper-administered versions of Bath ankylosing spondylitis (AS) questionnaires and the Quebec Scale for low back pain (LBP).

Methods. Fifty patients with LBP completed the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Functional Index (BASFI), Global Score (BAS-G), and the Quebec Scale. Outcome measures were administered both in traditional paper format and by computerized touch-screen system. The order of completion was randomly assigned to each participant. The length of time required to complete each set of questionnaires was recorded and a "washout" period of at least 40 minutes was ensured between completion of the first and second set of outcome measures.

Results. There was no statistically significant difference in completion time between the 2 methods of administration. A small systematic difference between computer and paper-administered versions was observed in the Quebec Scale and in the BAS-G results. However, there was a high degree of agreement between paper and computer-administered versions of the Quebec Scale, the BASDAI, BASFI, and BAS-G. Out of the 50 subjects, 84% indicated a preference for the computer-administered method.

Conclusion. The Bath AS questionnaires and the Quebec Scale can be reliably administered by a computerized touch-screen system. Given the ease of data integration and analysis supported by computer-administered versions of these outcome measures, their excellent reliability, and their popularity among study participants, the computerized versions of the BASDAI, BASFI, BAS-G, and Quebec Scale seem preferable to the traditional paper format. (J Rheumatol 2005;32:669–72)

Key Indexing Terms:
OUTCOME MEASUREMENT COMPUTER VALIDITY
SELF-ASSESSMENT ELECTRONIC QUESTIONNAIRES

Information technology has wide application in medicine. Computer-assisted administration of outcome measures is an example of a useful application in rheumatology. We designed a touch-screen computer system to allow outcome measures from standardized questionnaires to be directly entered into the computer database by the patient. The advantages of this system include reducing the use of paper, bypassing data entry normally required for transferring paper outcomes to computer database, and reduction in missing data resulting from skipped questions on the paper format. Outcome data are more complete and more easily reviewed and analyzed.

Few studies demonstrate the validity of using computerized systems to record outcome measures[1,2]. Expanding use of computerized technology will allow new and more efficient methods of data collection in clinical practice. However, it is important to establish validity and reliability before expanding the use of computerized formats.

We compared the validity and reliability of computer-administered versions of 2 common back pain outcome measures in rheumatology, the Quebec Scale and Bath ankylosing spondylitis (AS) outcome measures, with traditional paper-administered versions.

MATERIALS AND METHODS

Subject selection. Patients attending a tertiary referral rheumatology outpatient clinic were recruited consecutively. Patients reporting low back pain (LBP) were eligible. Patients unable to read or comprehend the questionnaires in English were excluded from the study.

Outcome measure administration. Upon arrival at the clinic, patients were invited to participate in the study. Consent was obtained in accord with ethics guidelines of the University of British Columbia/Providence Health Care Research Ethics Board. Upon consent, study subjects were asked to complete 4 outcome measures related to LBP: the Quebec Scale[3], the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)[4], the Bath

From St. Paul's Hospital Orthopaedic Medicine Clinic; and the Department of Health Care and Epidemiology, University of British Columbia, Vancouver, British Columbia, Canada.

Supported by the St. Paul's Hospital Foundation.

H. Bent, BA, PG Dip Psych; C.R. Ratzlaff, BSc (PT), FCAMT, St. Paul's Hospital Orthopaedic Medicine Clinic; E.C. Goligher, BSc, Med-1, Faculty of Medicine, University of British Columbia (UBC); J.A. Kopec, MD, PhD, Assistant Professor, Department of Health Care and Epidemiology, UBC; J.H. Gillies, MD, FRCPC, Clinical Instructor, St. Paul's Hospital Orthopaedic Medicine Clinic and Division of Rheumatology, UBC.

Address reprint requests to Dr. J.H. Gillies, Orthopaedic Medicine Clinic, St. Paul's Hospital, 401-1177 West Broadway, Vancouver, BC V6H 1G3, Canada. E-mail: jeangillies@telus.net

Accepted November 29, 2004.

Ankylosing Spondylitis Function Index (BASFI)[5], and the Bath Ankylosing Spondylitis Global score (BAS-G)[6]. The set of outcome measures were administered twice: by touch-screen computer system and by standard paper format. The order of the method of administration was randomly assigned to account for possible order effects.

A visual analog scale (VAS) was used to record the 3 Bath AS outcome measures: The BASDAI consists of six 10 cm VAS: 5 are classic VAS and the sixth is a variation of the VAS (the upper limit is not an absolute limit). Participants were instructed to answer the questions regarding perceived severity of fatigue, pain, swelling, or stiffness by making a mark along the 10 cm line. For the first 5 questions, the 10 cm line is anchored at one end by the label "None" and at the other end by "Very Severe"; for the sixth question, regarding length of morning stiffness, it is anchored at one end by "0 hours" and at the other end by "2 or more hours." The Global Score (BAS-G) consists of two 10 cm VAS relating to the participants' perception of well being. The scales have the same descriptors as the BASDAI. The BASFI consists of ten 10 cm VAS that require the subject to indicate their level of ability on a scale from 0 cm (easy) to a maximum of 10 cm (impossible). Participants were instructed to answer the questions by making a mark along the 10 cm line. The Quebec Scale measures disability arising from LBP. It comprises 20 questions, each a multiple choice of 6 options scored 0 (not difficult at all) to a maximum of 5 (unable to do).

The computerized touch-screen administration system allowed patients to record responses in a manner identical to standard paper format. The system administered each question of the outcome measures individually and consecutively. The subject was required to touch the computer screen at a point on the scale indicating the severity of their pain or disability. All VAS were exactly 10 cm in length (as measured on the computer screen) in accord with the published paper formats. The markings on the scales and the descriptions of the scale endpoints were also identical to the paper format. The wording of the questions of all the computerized outcome measures was identical to the wording of the questions in the paper format.

The length of time required to complete both the computer-administered and paper-administered set of questionnaires was recorded. There was a minimum 40 minute "washout" period between completion of the first set of outcome measures and commencing the second set. The subject had no access to prior scores when completing the second set of outcome measures.

Finally, subjects were asked to note whether they had any preference for either method of administration. Age, sex, and ethnic background of each subject were recorded.

Data treatment and analysis. The questionnaires were scored based on descriptions provided in the original articles reporting the Quebec Scale[3], BASDAI, BASFI[5], and BAS-G[6]. Systematic differences in responses between the 2 methods of administration were assessed by paired t test (parametric data) and Wilcoxon signed-rank test (nonparametric data). Agreement was measured by calculating the intraclass correlation coefficient (ICC) for each outcome measure. We used the ICC for a fixed effects 2-way analysis of variance[7]. We also employed the method described by Bland and Altman for evaluating agreement between 2 methods[8]. The difference in completion time and the order of administration were analyzed by the paired t test. Missing data points were resolved by substituting the average value of the questionnaire results for the missing data point. Statistical analysis was carried out using JMP 5.0.1 by the SAS Institute and S-Plus 6.1 by Insightful.

RESULTS

A total of 50 subjects were recruited for participation. The mean age of the study population was 49.5 years (interquartile range 41–61). Ninety-four percent of subjects were of Caucasian background and 6% were of Asian background. Seventy percent of subjects were female.

All subjects completed both questionnaires. Eighty-four percent reported a preference for the computerized version, 12% reported a preference for the paper version, and 4% reported no preference. The paper version of the outcome measures required 5 minutes and 47 seconds to complete on average, while the computer version required an average of exactly 6 minutes. The difference was not statistically significant (p = 0.68).

The order of test administration (i.e., taking paper test first vs computer test first) did not produce any significant difference in mean scores on the Quebec (p = 0.7301), BASDAI (p = 0.1312), BASFI (p = 0.8782), or BAS-G (p = 0.4348).

There was a small systematic difference between methods of administration in 2 of the scales. The results of the paired t test for each scale are shown in Table 1. It is interesting that the computer score was lower than the paper score on average, although only 2 of the differences were significantly different from zero.

Table 1. Systematic differences between methods of administration of questionnaires.

Subscale	n	Maximum Possible Score, cm	Mean Score Difference, cm (Computer–Paper)	95% CI
BASDAI	50	10	–0.07	–1.4, 1.27
BASFI	50	10	0.34	–2.26, 2.94
BAS-G	50	10	–0.86*	–1.6, –0.15
Quebec	50	100 points	–2.75* points	–4.47, –1.13

* Statistically significantly different from 0.

Table 2. Item-by-item differences between computer and paper-administered questionnaires.

Item	Mean Difference	p	Item	Mean Difference	p
BASDAI (Maximum score 10 cm)			Quebec (maximum score 5 pts)		
1	–0.10	0.66	1	–0.16	0.13
2	–0.47	0.12	2	–0.14	0.18
3	0.03	0.90	3	0.01	1.00
4	0.05	0.88	4	–0.28	0.08
5	0.04	0.86	5	–0.32	0.00*
6	0.22	0.33	6	–0.29	0.01*
BASFI (maximum score 10 cm)			7	–0.06	0.65
1	0.27	0.08	8	–0.11	0.12
2	0.05	0.83	9	–0.02	0.65
3	0.40	0.01*	10	–0.16	0.06
4	–0.17	0.58	11	–0.14	0.25
5	–0.23	0.40	12	0.02	0.99
6	–0.28	0.41	13	–0.17	0.20
7	0.01	0.95	14	–0.16	0.08
8	–0.05	0.84	15	–0.04	0.55
9	–0.10	0.79	16	–0.20	0.14
10	0.24	0.36	17	–0.12	0.24
BAS-G (maximum score 10 cm)			18	–0.12	0.23
1	–0.33	0.11	19	–0.12	0.30
2	–0.53	0.06	20	0.04	1.00

* Statistically significantly different from 0.

Table 3. Intermethod agreement.

Subscale	n	ICC
Quebec	50	0.95
BASDAI	50	0.94
BASFI	50	0.92
BAS-G	50	0.86

Item by item analysis was conducted using paired t tests (parametric data) and the Wilcoxon signed-rank test (nonparametric data) to compare scores on paper versus computer. Results are displayed in Table 2. Of the 38 items on the 4 scales, there were small significant differences for 3 items: one question on the BASFI and 2 on the Quebec scale.

Agreement between methods of administration was first assessed by the ICC. The results are displayed in Table 3. All the scales demonstrated strong agreement between the 2 methods. The ICC value for the BAS-G scale was lower, possibly because it contains fewer items[2].

Bland-Altman plots were prepared for the BASDAI, BASFI, BAS-G, and Quebec Scale (Figure 1). The limits of agreement between paper and computer versions for the Quebec Scale were about ± 10 points out of a possible 100. The BASDAI, BASFI, and BAS-G exhibited larger limits of agreement of about ± 20% between methods of administration. There was no observed relation between measurement error and observed score in any of the scales.

DISCUSSION

We investigated the validity and reliability of a computerized touch-screen system for the administration of common outcome measures for low back pain and ankylosing spondylitis in an outpatient rheumatology setting. We found good agreement between computerized and traditional formats of the questionnaires and a strong preference for the computerized method among study participants.

Our results indicate a small systematic tendency for the computer format of the outcome measures to score lower

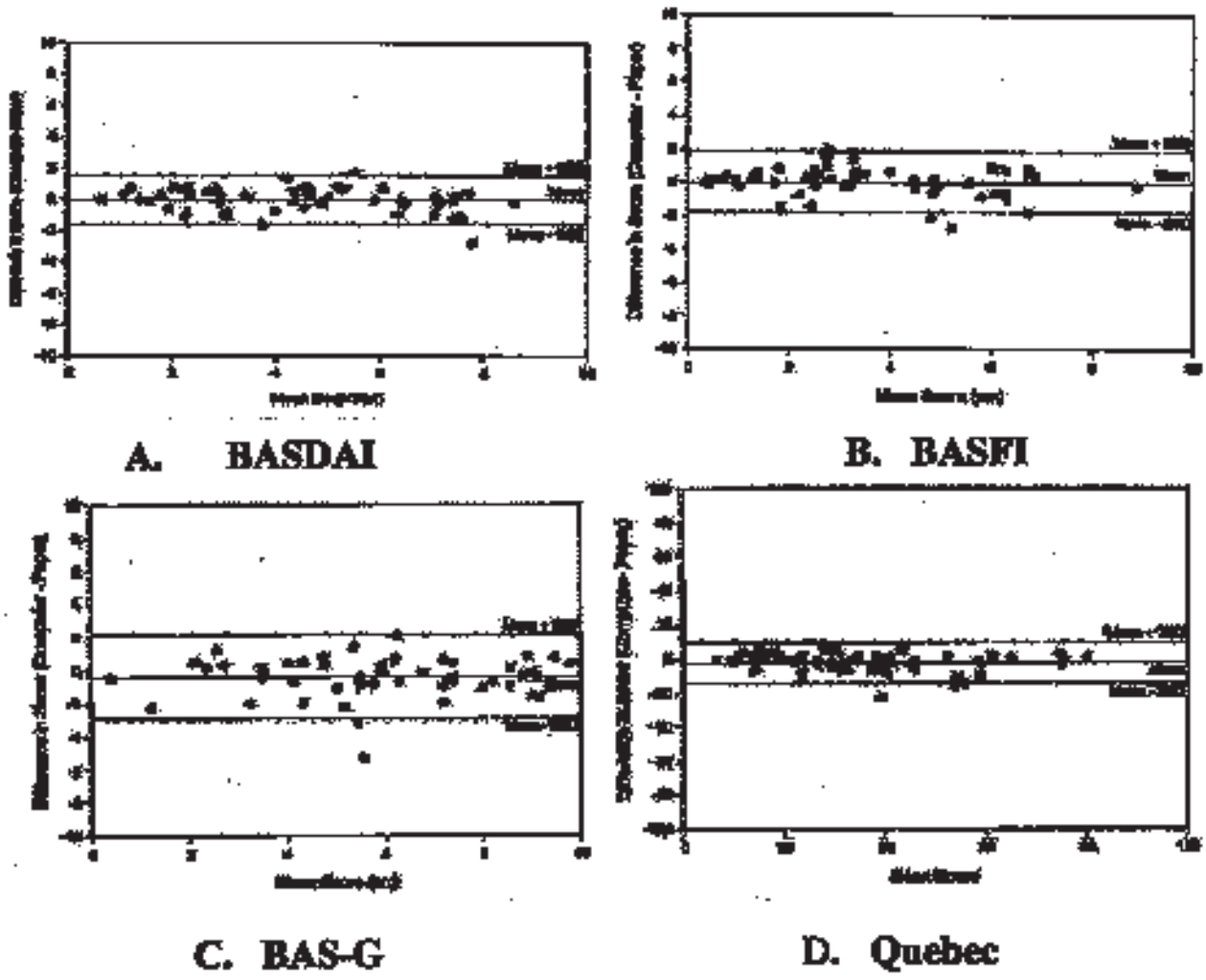

Figure 1. Bland-Altman plots of the difference between methods versus the mean: BASDAI (A), BASFI (B), BAS-G (C), and Quebec Scale (D).

than the paper format. Most of the individual item scores showed no significant differences between computer and paper-administered items. However, 2 individual items and 2 of the aggregate scale score comparisons showed a statistically significant decrease in the computer-administered results (one item showed an increase). A similar trend of slightly lower computer scores was observed by Bellamy and colleagues[1] in a study of a computerized version of the WOMAC (Western Ontario McMaster University Osteoarthritis Index). They speculated that the difference may be due to the vertical orientation of the computer screen versus the horizontal orientation of the paper or an altered perception of the VAS scale because of the differing interfaces. In our study the systematic differences between the aggregate scale scores were never more than 1–4% of the total scale score, a difference that we consider clinically insignificant. In the case of the VAS scales used by the Bath AS measures, it may be that the touch-screen system contributed to a small error in VAS scoring due to the relative thickness of a finger in selecting a point on the VAS scale in contrast to a pen marking a point on paper. This, however, would not explain the slightly lower score on the paper version of the Quebec Scale, as the questions require a categorical response. Irrespective of the cause, the small tendency for the computer format to score lower than the paper format may indicate that the same format should be used in any subsequent retesting with the same outcome measure.

One limitation of this study was that we are unable to differentiate between intermethod variability and random error. Test-retest variability of an outcome measure will contribute to increased variability between methods of administration of the outcome measure. However, some information on the test-retest reliability of these outcome measures is available in the literature. The Quebec Scale was reported to have a test-retest reliability slightly lower than the intermethod reliability we found in this study (ICC = 0.84)[9]. Test-retest reliability scores for the BASDAI, BASFI, and BAS-G were found to be r = 0.93[4], r = 0.89[5], and r = 0.84[6], respectively. While we did not use Pearson's r in our study, the ICC is generally found to be similar to (or slightly lower than) Pearson's r, so that our results are quite similar to these published values. This seems to indicate that a good proportion of observed intermethod variability in our study is due to random test-retest error.

A further potential limitation of this study was the relatively short washout period of 40 minutes between administration of the computer and paper versions. This may have allowed a memory effect to contribute to agreement between methods of administration. However, subjects did not have access to their scores from the first test when completing the second and, as 4 tests were administered in each version with a total of 38 questions, it was thought unlikely that the subjects would recall their scores from the first test. Bellamy, *et al*[1] in comparing paper and computer versions of the WOMAC VA3.0 used a 10 minute washout period. Further, in the population sampled for this study, scores over longer periods (e.g., several days) would likely have varied due to change in the variables being evaluated (i.e., pain, stiffness, etc).

During the analysis we observed that a number of the paper questionnaires had missing data. Evidently subjects did not complete them in full or their answers were so ambiguous as to render them meaningless. This problem points out a major advantage with the computerized method of administration; the computer software was designed so that subjects could not skip questions without providing an answer and the answer that they provided was always interpretable to the computer. This prevents loss of valuable data.

The findings of our study add to the growing body of evidence pointing to the usefulness, validity, and reliability of computerized outcome measures. Computers can increase the speed and efficiency of data collection and processing. In addition, our results indicate that patients may prefer entering answers into the computer. While previous studies have emphasized the usefulness of such technology for clinical trials, our study showed patients are able to use this technology in the outpatient rheumatology setting. We recommend the integration of computerized versions of the BASDAI, BASFI, BAS-G, and the Quebec scale into daily outpatient practice in rheumatology.

REFERENCES

1. Bellamy N, Campbell J, Stevens J, Pilch L, Stewart C, Mahmood Z. Validation study of a computerized version of the Western Ontario and McMaster Universities VA3.0 Osteoarthritis Index. J Rheumatol 1997;24:2413-5.
2. Theiler R, Spielberger J, Bischoff HA, Bellamy N, Huber J, Kroesen S. Clinical evaluation of the WOMAC 3.0 OA index in numeric rating scale format using a computerized touch screen version. Osteoarthritis Cartilage 2002;10:479-81.
3. Kopec JA, Esdaile JM, Abrahamowicz M, et al. The Quebec Back Pain Disability Scale. Spine 1995;20:341-52.
4. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21:2286-91.
5. Calin A, Garrett S, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281-5.
6. Jones SD, Steiner A, Garrett SL, Calin A. The Bath Ankylosing Spondylitis Patient Global Score (BAS-G). Br J Rheumatol 1996;35:66-71.
7. Armstrong BK, White E, Saracci R. Principles of exposure measurement in epidemiology. Oxford: Oxford University Press; 1994:102-3.
8. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1:307-10.
9. Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. Phys Ther 2002;82:8-24.