Performance of health status measures with a pen based personal digital assistant

T K Kvien, P Mowinckel, T Heiberg, K L Dammann, Ø Dale, G J Aanerud, T N Alme, T Uhlig

Background: Increasing use of self reported health status in clinical practice and research, as well as patient appreciation of monitoring fluctuations of health over time, suggest a need for more frequent collection of data. Electronic use of health status measures in the follow up of patients is a possible way to achieve this.

Objective: To compare self reported health status measures in a personal digital assistant (PDA) version and a paper/pencil version for test–retest reliability, agreement between scores, and feasibility.

Methods: 30 patients with stable rheumatoid arthritis (mean age 61.6 years, range 49.8 to 70.0; mean disease duration, 16.7 years; 63% female; 67% rheumatoid factor positive; 46.6% on disease modifying antirheumatic drugs) completed self reported health status measures (pain, fatigue, and global health on visual analogue scales (VAS), rheumatoid arthritis disease activity index, modified health assessment questionnaire, SF-36) in a conventional paper based questionnaire version and on a PDA (HP iPAQ, model h5450). Completion was repeated after five to seven days.

Results: Test–retest reliability was similar, as evaluated by the Bland–Altman approach, the coefficient of variation, and intraclass correlation coefficients. The scores showed acceptable agreement, but with a slight tendency to higher scores on VAS with the PDA than the paper/pencil version. No significant differences were seen for measures of feasibility (time to complete, satisfaction score), but 65.5% preferred PDA, 20.7% preferred paper, and 13.8% had no preference.

Conclusions: The clinimetric performance of paper/pencil versions of self reported health status measures was similar to an electronic version, using an inexpensive PDA.

Abbreviations: VAS, visual analogue scale; SF-36, short form 36 item health assessment questionnaire; OMERACT, outcome measures in rheumatoid arthritis; RADAI, rheumatoid arthritis disease activity index; SDD, smallest detectable difference; MHAQ, modified health assessment questionnaire; PDA, personal digital assistant.
questionnaire (MHAQ),9 and SF-36.20 A trained study nurse carried out joint counts on both occasions, and acute phase reactants were also examined to verify stability of the disease activity.

The patients were asked to rate the satisfaction with each method (PDA v paper/pencil) on a 100 mm VAS.

Feasibility was also assessed by the time needed to complete the self reported measures and by a final question asking for the preferred method. We also recorded whether patients were able actually to use the PDA.

Analysis of data

Recorded data in the PDA were transferred through a wireless local area network to a computer, and then imported as a text file into SPSS for analyses. The paper data were entered manually into the database. Test–retest reliability was examined by the Bland–Altman approach,21 computing the smallest detectable difference (SDD)—that is, 1.96 SD of the difference between the scores, as well as the limits of agreement,21 by the coefficient of variation (CV%), and by intraclass correlation coefficients (ICC). We used a two way mixed effects model to calculate the single score ICC. The reliability of the PDA and paper versions was compared by an analysis of variance (ANOVA) on the differences. Relations between the scores were examined by Pearson correlation coefficients and agreement by the computed differences between the scores with a 95% confidence interval (CI) as well as the limits of agreement,21 by the coefficient of variation (CV%), and by the difference between the scores, as well as the limits of agreement,21 by the coefficient of variation (CV%), and by intraclass correlation coefficients (ICC). We used a two way mixed effects model to calculate the single score ICC. The reliability of the PDA and paper versions was compared by an analysis of variance (ANOVA) on the differences. Relations between the scores were examined by Pearson correlation coefficients and agreement by the computed differences between the scores, as well as the limits of agreement,21 by the coefficient of variation (CV%), and by intraclass correlation coefficients (ICC). We used a two way mixed effects model to calculate the single score ICC. The reliability of the PDA and paper versions was compared by an analysis of variance (ANOVA) on the differences. Relations between the scores were examined by Pearson correlation coefficients and agreement by the computed differences between the scores, as well as the limits of agreement,21 by the coefficient of variation (CV%), and by intraclass correlation coefficients (ICC).

The patients were able actually to use the PDA.

RESULTS

Thirty patients were included after giving their informed consent (table 1). The SDDs, limits of agreement, CV%, and ICCs are shown in table 2 (data for SF-36 role scales and social functioning are not included). The SDDs were rather high, but similar for paper and PDA. The ICC varied between 0.69 and 0.93 for the paper versions and between 0.66 and 0.94 for the PDA versions. The results of analysis of variance confirmed that the reliability was similar with the two methods (p>0.10 for all measures, data not shown). Measures of reliability were of the same magnitude even when replacing one method by the other at T1 and T2 (for example, examining scores by paper at T1 and PDA at T2).

The scores obtained with the two methods correlated strongly with each other (r values between 0.79 and 0.97 at T1 and 0.82 to 0.98 at T2) (table 3). Higher VAS scores were seen with the paper than with the PDA versions in four of the six measures (p = 0.03 to 0.05), but the statistical significance disappeared after Bonferroni correction for the number of tests. No differences between the methods were observed for the scale scores (MHAQ, RADAI, and SF-36) except for a difference for SF-36 bodily pain at T1 and mental at T2, with more severe scores for the PDA than for the paper version (table 3). The agreement between scores obtained by the PDA and paper versions is also illustrated by Bland–Altman plots for RADAI and pain VAS at T1 in fig 1A and 1B.

All patients managed to use the PDA. The mean (SD) satisfaction score at T1 with paper was 67.9 (18.1) and with PDA, 60.4 (11.4) (p = 0.29). Mean (SD) time to complete was 24.9 (27.0) and 30.5 (16.0) minutes, respectively (p = 0.11). Similar results were seen at T2. At the end, 20.7% reported preference for the paper version, 65.5% for the PDA, and 13.8% had no preference.

DISCUSSION

This study showed that widely used health status measures performed similarly in a format for PDA and in the traditional paper/pencil format. The recent OMERACT–7 meeting recommended that electronic data collection techniques for outcome assessment should be adequately validated.29 It is possible that electronic and paper versions of questionnaires may provide different results, taking into account that patients with a PDA are presented with one question at a time, and then have to proceed to the next screen to see the next question. Thus there are 36 screens for SF-36, 13 for RADAI, and one screen for each VAS. In

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Demographic characteristics and baseline scores of disease activity measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>61.6 (49.8 to 70.0)</td>
</tr>
<tr>
<td>Female</td>
<td>63%</td>
</tr>
<tr>
<td>Disease duration (years)</td>
<td>16.7 (3.5 to 45.4)</td>
</tr>
<tr>
<td>Erosive disease</td>
<td>72%</td>
</tr>
<tr>
<td>Rheumatoid factor</td>
<td>67%</td>
</tr>
<tr>
<td>28-SJC</td>
<td>1.9 (0 to 10)</td>
</tr>
<tr>
<td>28-TJC</td>
<td>3.4 (0 to 16)</td>
</tr>
<tr>
<td>ESR (mm/h)</td>
<td>15 (3 to 52)</td>
</tr>
<tr>
<td>C reactive protein (mg/l)</td>
<td>8.8 (2 to 75)</td>
</tr>
</tbody>
</table>

Values are mean (range) for continuous variables and per cent for counts.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Mean scores at baseline, and test–retest reliability of patient reported measures in paper and personal digital assistant (PDA) format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper format</td>
<td>PDA format</td>
</tr>
<tr>
<td>Score T1</td>
<td>SDD</td>
</tr>
<tr>
<td>Pain VAS</td>
<td>28.8</td>
</tr>
<tr>
<td>Fatigue VAS</td>
<td>42.4</td>
</tr>
<tr>
<td>Patient global VAS</td>
<td>29.2</td>
</tr>
<tr>
<td>RADAI</td>
<td>3.16</td>
</tr>
<tr>
<td>MHAQ</td>
<td>1.51</td>
</tr>
<tr>
<td>SF-36</td>
<td></td>
</tr>
<tr>
<td>Physical</td>
<td>60.2</td>
</tr>
<tr>
<td>Mental</td>
<td>74.7</td>
</tr>
<tr>
<td>Pain</td>
<td>56.0</td>
</tr>
<tr>
<td>Vitality</td>
<td>47.9</td>
</tr>
<tr>
<td>General</td>
<td>50.3</td>
</tr>
</tbody>
</table>

CV, coefficient of variation; ICC, intraclass correlation coefficient; LA, limits of agreement (that is, 95% confidence interval of the difference T1 minus T2); MHAQ, modified health assessment questionnaire; RADAI, rheumatoid arthritis disease activity index; SDD, smallest detectable difference (that is, 1.96 SD of the difference T1 minus T2); SF-36, short form 36 item health assessment questionnaire; VAS, visual analogue scale.
focusing on VAS \(^2\) and in patients with acute \(^1\) and chronic pain, \(^2\) orthopaedic conditions, \(^2\) gastro-oesophageal reflux, \(^2\) and Parkinson’s disease. \(^2\) Scores that were obtained with PDA and paper in the present study were also similar (table 3), but the Bland–Altman plots indicated that some individuals had major differences between their scores (fig 1A and 1B).

The test–retest reliability of patient reported measures was satisfactory when considering the ICCs. The 95% SDDs provide clinically useful information, as they represent the cut off values that have to be exceeded if a clinician wants to be 95% confident that a change reflects a significant improvement or deterioration. These values were 0.16 with paper and 0.20 with PDA for MHAQ, 1.11, and 1.24 for RADAI, and 14.8 and 12.7 for SF-36 physical, respectively (table 2). The value for MHAQ was close to the change that has been considered clinically important with the health assessment questionnaire (HAQ). \(^3\) We also repeated the joint counts and the acute phase reactants at T2. For comparison the SDD of the 28 swollen joint count (28-SJC) was 2.8, for the 28 tender joint count (28-TJC) it was 3.48, for investigators global it was 8.0, and for ESR/C reactive protein it was 5.7/4.8 (data not shown). The magnitude of the measurement errors is important when clinical decisions are based on changes in scores in individual patients. These measurement errors also suggest that frequent monitoring may be relevant for clinical decision making in individual patients, as repeated measures may provide more reliable information than, for example, single assessments at scheduled clinic visits. One feasible approach to obtaining such frequent data is electronic recording of patient reported data \(^2\) with wireless transferral of data directly into the hospital computer system.

Recall of a previous recording of health may influence the second recording. We assume that this problem is reduced when large numbers of instruments are completed as in this study. The completion of the instruments should ideally have been done in a randomised crossover design. The patients in this study were for practical reasons examined in four groups of six to eight individuals, and all patients within one group completed all instruments in the same sequence. However, two of the groups started with PDA and the other two started with the paper/pencil version.

Individual patients may differ with regard to acceptance of new technology. We chose to carry out this study in a typical age span of rheumatoid arthritis—that is, 50 to 70 years—and not in younger patients where acceptance could be expected to be high. This study was not powered to explore the performance of the electronic version in subgroups of patients based on sex, level of education, or age. The impression from the investigators was that no particular subgroup preferred one method to the other, and studies in other patient groups have not revealed any major problems in subgroups based on computer literacy, educational level, age, sex, or race. \(^4\) \(^2\) \(^2\) We had expected that feasibility would be higher with paper than with PDA, taking into account that we recruited regular patients without any particular bias in the direction of new technology. Time to complete was numerically lower with paper than with PDA. We also had experiences during the study that may lead to improved feasibility. For example, some dissatisfaction was related to slowness in the electronic wireless transferral of data from the PDA to the computer.

Feasibility for researchers obviously favours electronic recording, as the data may be fed directly into the computer system without any data entry procedures, which may prevent incorrect data being entered and reduce costs. Compliance with data recording might be improved because patients are prompted to answer questions at the right time.
moment—that is, they have to complete one screen before they are allowed to proceed to the next.\textsuperscript{1\textminus 26} Furthermore, the time and date of the completion are also recorded.

Electronic versions of patient reported health status have major potential for the future. With increasing demands for documentation of results of expensive drugs and other interventions in health care, patients may do their self assessments at home in the near future. Different types of technology can be used. The advantage of PDA is cost and flexibility (the device can be carried and used everywhere). However, access to home based computers or laptops will increase in the future, and computers with touch screens for data recording in the hospital represents another opportunity. Williams et al\textsuperscript{14} have described the positive and cost saving experiences with a series of assessments from questionnaire in a setting of 1062 patient visits. Data collected by computers can be accessed immediately on the clinician’s computer screen if the data system includes software that can compute scores (for example, MHAQ, SF-36, RADAI) from the data that have been entered. Mistakes during data entry can be prevented by a warning if values outside predefined ranges are entered, and costs related to data entry from paper questionnaires are saved. Data entered in patients’ homes can also be transferred to the hospital network through a cell phone network (in the case of PDA) or through the internet (in the case of a personal computer).

The potential is probably also large in clinical trials. Frequent assessments may be mandatory for some interventions—for example, treatment of acute pain—and responsiveness may be improved if repeated measures are available instead of a single measure from one time point. Validation of responsiveness was not part of this study but it is one of the next steps in the additional validation of health status reported on a PDA in patients with rheumatoid arthritis.

Are there any potential disadvantages? Some patients may have barriers to the use of new technology,\textsuperscript{14} even though this was not experienced in the present investigation and has not been a problem in other settings.\textsuperscript{26} 2\textsuperscript{6} Security systems with firewalls around the hospital server are mandatory to secure confidentiality and privacy. Further, patients with severe or worsening hand problems may not be able to use the small stylus to enter data in a PDA. This may cause selective drop out from the study of patients with worsening hand function.

In this project we were particularly aware of this potential problem, as also one of the members of the research group (OD) is an occupational therapist. However, our experiences indicate that almost all patients can use PDAs, and some patients with finger deformities replace the stylus with the fingertip on the touch screen of the PDA or a thicker pen. We did not observe problems with visibility of the screen, and all patients managed to use the PDA.

The performance of paper based versions of self reported health status measures and electronic versions, using an inexpensive PDA, was similar in this study. In our opinion, both regular computers and PDA have future potential for monitoring and assessing disease. Advantages of PDA are the size and availability at virtually any time and place, provided there is access to power to charge the batteries. Our results encourage validation of electronic recording of self reported health status in different clinical and research settings in rheumatology, and especially further research on the responsiveness and the daily recording of health status in the patients’ homes.

ACKNOWLEDGEMENTS

We are grateful for financial support from Grethe Harbitz’s Legacy and the Norwegian Rheumatism Association.

Authors’ affiliations

TK Kvien, Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway

P Mowinkel, K L Dammann, Ø Dale, G J Aanerud, T Uhlig, National Resource Centre for Rehabilitation in Rheumatology, Diakonhjemmet Hospital

T Heiberg, Department of Administration, Ulleval University Hospital, Oslo

T N Alme, Medicom, Oslo

REFERENCES


Table 3 Correlation (\(r\), Pearson correlation coefficients) between patient reported measures in paper and personal digital assistant (PDA) format, and difference PDA minus paper

<table>
<thead>
<tr>
<th>T1</th>
<th>Difference</th>
<th>95% CI</th>
<th>T2</th>
<th>Difference</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain VAS</td>
<td>0.87</td>
<td>4.5</td>
<td>0.1 to 8.9</td>
<td>0.93</td>
<td>3.6</td>
</tr>
<tr>
<td>Fatigue VAS</td>
<td>0.89</td>
<td>-2.1</td>
<td>-7.6 to 3.3</td>
<td>0.93</td>
<td>3.8</td>
</tr>
<tr>
<td>Patient global VAS</td>
<td>0.86</td>
<td>3.8</td>
<td>-0.5 to 8.1</td>
<td>0.84</td>
<td>4.8</td>
</tr>
<tr>
<td>MHAQ</td>
<td>0.93</td>
<td>-0.1</td>
<td>-0.3 to 0.2</td>
<td>0.98</td>
<td>0.0</td>
</tr>
<tr>
<td>SF-36</td>
<td>0.97</td>
<td>-0.01</td>
<td>-0.05 to 0.03</td>
<td>0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>Physical</td>
<td>0.94</td>
<td>-0.9</td>
<td>-4.1 to 2.2</td>
<td>0.95</td>
<td>2.1</td>
</tr>
<tr>
<td>Mental</td>
<td>0.91</td>
<td>3.7</td>
<td>-7.1 to 0.4</td>
<td>0.95</td>
<td>-4.3</td>
</tr>
<tr>
<td>Pain</td>
<td>0.79</td>
<td>-6.0</td>
<td>-11.4 to -0.6</td>
<td>0.82</td>
<td>0.1</td>
</tr>
<tr>
<td>Vitality</td>
<td>0.86</td>
<td>0.7</td>
<td>-3.6 to 5.1</td>
<td>0.89</td>
<td>-0.7</td>
</tr>
<tr>
<td>General</td>
<td>0.84</td>
<td>4.1</td>
<td>-1.5 to 9.7</td>
<td>0.82</td>
<td>5.1</td>
</tr>
</tbody>
</table>

Value are mean (95% confidence interval). CI, confidence interval; MHAQ, modified health assessment questionnaire; RADAI, rheumatoid arthritis disease activity index; SF-36, short form 36 item health assessment questionnaire; VAS, visual analogue scale.