

available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/rmed

Patients may respond differently to paper and electronic versions of the same questionnaires

Elizabeth F. Juniper^{a,*}, John M. Langlands^b, Bridget A. Juniper^c

^a Department of Clinical Epidemiology and Biostatistics, McMaster University, Ontario, Canada

^b Langlands and Associates Consulting, Richmond, British Columbia, Canada

^c QOL Technologies Ltd, Bosham, West Sussex, England, UK

Received 30 April 2008; accepted 23 October 2008

KEYWORDS

Asthma;
Electronic;
Quality of life;
Questionnaire;
Rhinoconjunctivitis

Summary

Aim: To adapt the Asthma Quality of Life Questionnaire (AQLQ(S)), the Asthma Control Questionnaire (ACQ) and the Rhinoconjunctivitis Quality of Life Questionnaire (RQLQ(S)) for a personal digital assistant (Palm TX) and to examine the validity of the electronic versions by comparing them with the original paper versions.

Methods: 84 adults with asthma and 32 with rhinitis were randomised to complete either the paper or the electronic version first. After 2 h, they completed the other version.

Results: 68 asthma and 27 rhinitis patients provided analysable data. For the AQLQ(S) and RQLQ(S) differences between paper and electronic were significant. Concordance between paper and electronic, evaluated using an intraclass correlation coefficient were: AQLQ = 0.92, ACQ = 0.90 and RQLQ = 0.85. Concordance for the individual domains of the AQLQ and RQLQ ranged from 0.52 to 0.94. These levels of concordance did not reach the *a priori* defined requirement for validity.

Conclusions: The significant bias between paper and electronic versions and only modest concordance provides evidence that patients may respond differently to questionnaires in different formats and show that different formats must not be used interchangeably.

© 2008 Elsevier Ltd. All rights reserved.

Introduction

The 32-item standardised Asthma Quality of Life Questionnaire (AQLQ(S))^{1,2} and the 28-item standardised

Rhinoconjunctivitis Quality of Life Questionnaire (RQLQ(S))^{3,4} were developed to measure the functional problems that patients with asthma and rhinoconjunctivitis experience in their daily lives. The 7-item Asthma Control Questionnaire (ACQ)⁵ was developed to measure the primary clinical goal of asthma management. Patients respond to each question using a 7-point scale. All three questionnaires were developed in paper format in which medium they have undergone extensive validation both in

* Corresponding author. 20 Marcuse Fields, Bosham, West Sussex PO18 8NA, England, UK. Tel.: +44 (0) 1243 572124; fax: +44 (0) 1243 573680.

E-mail address: juniper@qoltech.co.uk (E.F. Juniper).

English and other languages. All three questionnaires have shown strong and consistent measurement properties and validity.

Many clinicians, academics and commercial companies are changing from paper to electronic data collection. A wide range of devices is being used including personal computers, the Internet, personal digital assistants (PDAs), scanners, interactive voice response (IVR), etc. Reasons for electronic data collection include reduced transcription errors, timed data entry, reduced risk of data loss, reduced storage space, reduced risk of data tampering and data collection away from the clinic. However, the validity of questionnaires may be specific to the method of data collection. It cannot be assumed that questionnaires validated using pen and paper are similarly valid when administered using alternate data collection methods.

In this study, the AQLQ(S), the ACQ and the RQLQ(S) were adapted for a PDA, the Palm TX. Validity was examined by comparing patients' responses to the original paper and PDA versions.

Materials and methods

Development of the electronic versions

The small screens of PDAs cannot contain a complete question plus a 7-point response scale. In most cases a single question and its responses could be fitted on two consecutive screens but for longer questions, three screens were required. Screens were carefully formatted and particular care was taken when rewording was necessary. Care was also taken to maintain the conceptual accuracy of the questions and the interval properties of the 7-point scales. All three questionnaires underwent a cognitive debriefing in a wide range of patients to ensure ease of use and accuracy of understanding of the instructions, questions and response options. The final English version was adapted for Polish, Russian, Ukrainian, Bulgarian and Czech using the linguistically validated paper versions in these languages. All new wording was translated and carefully checked by the MAPI Research Institute (Lyon, France) who had done the original linguistic validation for each language.

Validation

84 adults (17–65 years) with asthma participated in the AQLQ(S) and ACQ comparison. 32 of these patients also had rhinoconjunctivitis and were enrolled in the RQLQ(S) comparison. They were enrolled from patients participating in a clinical trial conducted by Inflazyme Pharmaceuticals (Richmond, British Columbia, Canada). For our study, patients with asthma were required to have AQLQ(S) <6.0 and ACQ >1.5. Patients with rhinitis were required to have RQLQ(S) >1.5. All patients gave informed consent.

Paper and electronic versions were completed during a single clinic visit with a two-hour interval in between. The AQLQ(S) and the ACQ were completed during the clinical trial screening visit and the RQLQ(S) comparison was done 2 weeks later. Patients were randomised to complete either the paper or the electronic version first. During the two-

hour interval between completions, patients carried out a non-medical activity (e.g. read, TV, etc.)

Statistical analysis

Overall and domain scores for electronic and paper versions were compared using a paired *t*-test. Concordance between the two methods was examined with an intraclass correlation coefficient (ICC). It was decided *a priori* that concordance between electronic and paper for the AQLQ(S) and the RQLQ(S) would be acceptable if the ICC for the overall score was 0.95 or greater. Concordance for the ACQ was required to be 0.90 or higher. These values are slightly lower than the test–retest reproducibility of the paper versions completed one week apart^{1–5} and in keeping with levels of concordance observed in previous validation studies of electronic devices.^{6,7} The ACQ was analysed both with and without FEV₁ data.

Results

Of the 84 patients with asthma who were randomised, 66 patients completed the AQLQ(S) and 68 completed the ACQ in both paper and electronic format and were included in the analysis. Three subjects were excluded because they completed the paper version in Ukrainian and the electronic in Russian. 16 were excluded from the AQLQ(S) and 14 from the ACQ analysis because they completed the electronic version of the questionnaires twice due to a technical fault with the device. 27 patients completed both paper and electronic versions of the RQLQ(S) and were included in the analysis. Six patients were excluded for missing electronic data and one was excluded for missing paper data.

Overall and domain scores, paired *t*-test results and concordance (intraclass correlation coefficients) for all three questionnaires are shown in Table 1. For the AQLQ(S), differences were significant for the overall score ($p = 0.009$) and two of the four domains ($p = 0.007$, $p = 0.01$). Concordance was ICC = 0.92 which was below the 0.95 level required for validity. For the ACQ, the difference did not reach statistical significance ($p = 0.12$) and concordance was ICC = 0.90. Differences for the RQLQ(S) were the borderline for the overall score ($p = 0.05$) and significant for 2 of the 7 domains. Concordance for the overall score was ICC = 0.84 which was well below the 0.95 level required for validity.

Discussion

The results of this study suggest that even with very careful formatting, rewording, linguistic validation and a thorough cognitive debriefing in patients, questionnaires adapted for another medium may not give data that is consistent with the original validated version. The study shows that these versions of the AQLQ(S) and RQLQ(S) on to the Palm TX gave inadequate concordance between paper (gold standard) and PDA to support the validity of these PDA versions. Of serious concern was the bias shown for the overall AQLQ(S) and RQLQ(S) scores and some of their domains. For the AQLQ(S), the difference was highly significant ($p = 0.009$).

Table 1 Comparison between electronic and paper version of the AQLQ(S), ACQ and RQLQ(S).

	Electronic mean $\pm$ sd	Paper mean $\pm$ sd	Difference	paired <i>t</i> -test <i>p</i> value	Pearson correlation coefficient	Intraclass correlation coefficient
AQLQ(S) overall	3.65 $\pm$ 0.72	3.75 $\pm$ 0.76	0.10	0.009	0.92	0.92
Symptoms	3.54 $\pm$ 0.72	3.60 $\pm$ 0.77	0.06	0.14	0.90	0.90
Activities	3.89 $\pm$ 0.79	4.01 $\pm$ 0.81	0.12	0.007	0.90	0.89
Emotions	3.51 $\pm$ 1.09	3.63 $\pm$ 1.03	0.12	0.10	0.86	0.85
Environment	3.50 $\pm$ 0.95	3.60 $\pm$ 0.95	0.11	0.01	0.94	0.94
ACQ	3.04 $\pm$ 0.52	2.99 $\pm$ 0.56	0.05	0.12	0.90	0.90
ACQ (no FEV ₁)	2.91 $\pm$ 0.58	2.86 $\pm$ 0.64	0.05	0.12	0.91	0.90
RQLQ overall	2.79 $\pm$ 0.89	2.59 $\pm$ 0.99	0.20	0.05	0.87	0.84
Activities	3.04 $\pm$ 0.98	3.01 $\pm$ 1.05	0.02	0.83	0.83	0.83
Sleep	2.77 $\pm$ 1.30	2.75 $\pm$ 1.30	0.01	0.96	0.51	0.52
Non-rhinitis symptoms	2.74 $\pm$ 1.10	2.47 $\pm$ 1.21	0.28	0.04	0.84	0.82
Practical problems	3.25 $\pm$ 1.23	3.02 $\pm$ 1.60	0.24	0.15	0.90	0.86
Nasal symptoms	2.95 $\pm$ 1.16	2.76 $\pm$ 1.31	0.19	0.19	0.82	0.81
Eye symptoms	2.24 $\pm$ 1.22	2.01 $\pm$ 1.23	0.23	0.15	0.79	0.78
Emotional function	2.81 $\pm$ 1.21	2.45 $\pm$ 1.50	0.36	0.07	0.75	0.71

The difference in RQLQ(S) overall score was borderline ($p = 0.05$) but with only 27 patients, the probability of a type 2 error was high and the bias probably real. Only the ACQ reached the required concordance for validity and even this was borderline.

The reason for the bias between paper and electronic is unclear. The order in which patients completed the two versions was randomised and therefore bias cannot be attributed to a learning effect or other external confounders. It may be that having questions and responses on different screens or in a changed format influenced how patients responded. On a different PDA (Palm Treo 650), the overall RQLQ(S) reached the required concordance with an ICC of 0.95 but it too showed significant bias in some of the domains.⁸ In contrast, large screen electronic versions of these questionnaires, where instructions, questions and their response options are on a single screen, concordance between paper and electronic was high (ICC = 0.99 and 0.96) with no evidence of bias.^{6–8}

The reason for the poor concordance between paper and electronic in this study, compared with a previous PDA study⁸ is also unclear. One possible reason is that the electronic configuration and software was not the same in the two devices. Secondly, this study was appended to a large multinational clinical trial and attention to protocol adherence may not have been as good.

Of additional concern was the amount of electronic data 'lost' by the PDAs (RQLQ(S): $n = 6$) and double entry of data due to technical faults (AQLQ(S): $n = 16$ and ACQ: $n = 14$). Although data integrity is marketed as an advantage of electronic data capture, this study shows that data losses do occur.

The results of this study and those from IVR validation studies,⁹ where significant biases were also observed, emphasise the importance of checking the validity of all new formats of questionnaires before they are used either in clinical practice or research. In addition, the results show that different formats of the same questionnaire cannot be used interchangeably.

Conflict of interest

None of the authors have any conflict of interest with any of the issues associated with this manuscript.

Acknowledgements

We thank Inflazyme Pharmaceuticals for funding this study and Alan Shields of PRO Consulting for conducting the statistical analysis.

References

1. Juniper EF, Guyatt GH, Ferrie PJ, Griffith LE. Measuring quality of life in asthma. *Am Rev Respir Dis* 1993;147:832–8.
2. Juniper EF, Buist AS, Cox FM, Ferrie PJ, King DR. Validation of a standardized version of the asthma quality of life questionnaire. *Chest* 1999;115:1265–70.
3. Juniper EF, O'Byrne PM, Guyatt GH, Ferrie PJ, King DR. Development and validation of a questionnaire to measure asthma control. *Eur Respir J* 1999;14:902–7.
4. Juniper EF, Guyatt GH. Development and testing of a new measure of health status for clinical trials in rhinoconjunctivitis. *Clin Exp Allergy* 1991;21:77–83.
5. Juniper EF, Thompson AK, Ferrie PJ, Roberts JN. Validation of a standardized version of the rhinoconjunctivitis quality of life questionnaire. *J Allergy Clin Immunol* 1999;104:364–9.
6. Caro JJ, Caro I, Caro J, Wouters F, Juniper EF. Does electronic implementation of questionnaires used in asthma alter responses compared to paper implementation. *Qual Life Res* 2001;10:683–91.
7. Bushnell DM, Martin ML, Parasuraman B. Electronic versus paper questionnaires: a further comparison in persons with asthma. *J Allergy* 2003;40:1–11.
8. Juniper EF, Riis B, Juniper BA. Development and validation of an electronic version of the rhinoconjunctivitis quality of life questionnaire. *Allergy* 2007;62:1091–3.
9. Juniper EF, Schatz M, Juniper BA, Zeiger RS. Evaluation of two interactive voice response telephone versions of health-related quality of life questionnaires. *J Allergy Clin Immunol* 2008;122:654–5.