

Demonstrating Measurement Equivalence of the Electronic and Paper Formats of the Urticaria Patient Daily Diary in Patients with Chronic Idiopathic Urticaria

Emuella M. Flood · James L. Zazzali ·
Jennifer Devlen

Published online: 6 August 2013
© Springer International Publishing Switzerland 2013

Abstract

Background The Urticaria Patient Daily Diary (UPDD), originally developed on paper, is a measure of key symptoms of chronic idiopathic urticaria (CIU). The development of the electronic version (eUPDD) involved moderate modifications to the appearance of the paper version.

Objective This study assessed the measurement equivalence of the electronic and paper versions of the UPDD in a sample of patients with CIU.

Methods This was a cross-over study of patients with moderate–severe CIU refractory to H1 antihistamines. Patients were randomized to either the eUPDD followed by the paper UPDD or vice versa. The UPDD includes morning and evening questions; both sets were administered together in this study. An hour-long filler task was given between paper and electronic administrations. Patients with stable symptoms between the two assessments were included in the analyses. Cohen's kappa coefficients and intraclass correlation coefficients (ICC) were computed as applicable to assess equivalence.

Electronic supplementary material The online version of this article (doi:10.1007/s40271-013-0021-4) contains supplementary material, which is available to authorized users.

E. M. Flood (✉) · J. Devlen
Oxford Outcomes, Inc., An ICON plc Company,
7315 Wisconsin Ave., Suite 250W, Bethesda,
MD 20814, USA
e-mail: emuella.flood@oxfordoutcomes.com

J. Devlen
e-mail: jennifer.devlen@oxfordoutcomes.com

J. L. Zazzali
Genentech, Inc., 1 DNA Way, South San Francisco,
CA, USA
e-mail: zazzali.james@gene.com

Results A total of 91 patients participated (mean age 43 years, 79.1 % female). Symptoms were stable between assessments for 67–74 (74–81 %) patients (varied by symptom). Kappa coefficients ranged from 0.82 to 1.00 for the individual UPDD items. For the Urticaria Activity Score (the sum of the 'itch severity' and 'number of hives' item scores) the ICC was 0.90 for the morning (Wilcoxon $p = 0.331$) and 0.95 for the evening (Wilcoxon $p = 0.836$).

Conclusions All test–retest statistics in this study were well above the accepted threshold, indicating excellent agreement between the two administration methods. Findings support the measurement equivalence of the electronic and paper versions of the UPDD to measure CIU symptoms.

Key Points for Decision Makers

- Electronic versions of patient-reported outcome measures should be tested for equivalence to validated paper versions of the measures.
- In this study, a diverse sample of patients with chronic idiopathic urticaria completed both electronic and paper versions of the Urticaria Patient Daily Diary during one 90-min study visit.
- Item-level analyses provide strong evidence to support the equivalence of the electronic and paper versions of the Urticaria Patient Daily Diary.

1 Background

Chronic idiopathic urticaria (CIU) or, as it is referred to in Europe, chronic spontaneous urticaria (CSU), is a relatively common skin disorder that affects approximately one in five individuals in their lifetimes, with an estimated point prevalence of 0.6 % of the population [1]. The disorder is characterized by repeated occurrences, over a

period of more than 6 weeks, of red, itchy wheals (or hives) and/or angioedema. Individual hives last for less than 24 h and can be as large as several centimeters in diameter. Itch, the hallmark symptom of CIU, typically varies throughout the day, with increased severity during the evening and night-time hours [2].

As urticaria symptoms tend to vary from day to day and even hour to hour, it is advised that patients track their symptoms at least once every 24 h in order to accurately describe their disease activity to their treating physicians. The European Academy of Allergy and Clinical Immunology has proposed using the Urticaria Activity Score (UAS) as the official disease activity measure. The UAS is a composite measure consisting of two scores—one score for the number of wheals (hives) and one score for the pruritis (itch severity)—which are summed for an overall score that ranges from 0 to 6 [3]. Additionally, formal testing has found that the UAS is a valid tool for the measurement of chronic urticaria symptoms [4]. Several recent clinical trials have utilized the UAS as the primary endpoint for assessing treatment efficacy. In 2009, Wan [5] successfully used the UAS as the primary outcome measure in a sample of 120 chronic idiopathic urticaria patients treated with a leukotriene receptor antagonist and an antihistamine. Kaplan et al. [6] and Serhat Inaloz et al. [7] combined the UAS score with the size of the largest hive for application in chronic urticaria patients in trials of omalizumab and cyclosporin, respectively.

Although several different measures of urticaria disease activity have been developed, the published urticaria scales remain in paper format. One such measure that includes the UAS is the Urticaria Patient Daily Diary (UPDD) [8]. The UPDD was developed in accordance with US FDA guidance for patient-reported outcome (PRO) measures [9], beginning with concept elicitation interviews that led to the creation of the diary items and included cognitive debriefing interviews in adults and adolescents with CIU, ensuring all items and instructions for the diary were comprehensive and clear [8, 10]. It has been shown to be a psychometrically reliable, valid, and responsive instrument in a sample of adults and adolescents aged 13–70 years, with moderate to severe CIU [11].

The development of the electronic version of the UPDD, the eUPDD, provides the opportunity to collect key urticarial symptom data more efficiently and conveniently. However, when a paper PRO measure is modified to be administered electronically, it is important to ensure that its validity is not undermined by the modifications. Thus, the current study was conducted to assess the equivalence between the paper UPDD and the eUPDD to ensure that patients interpret and respond to items in the same way regardless of format.

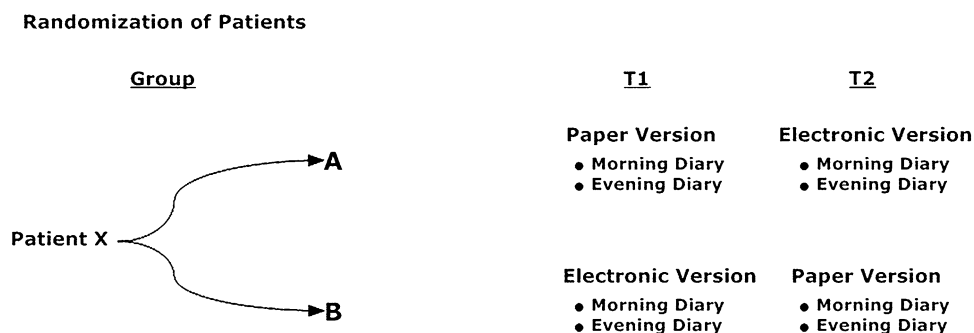
2 Methods

This study used a randomized cross-over design to assess the equivalence of paper and electronic versions of the UPDD at 11 clinical centers in the USA. Subjects were required to (i) have a diagnosis of moderate to severe CIU refractory to H1 antihistamines with the presence of pruritus and hives for more than 4 days per week for more than 6 weeks; (ii) be at least 12 years of age; and (iii) have adequate written and oral fluency in English. Exclusion criteria included presence of an inflammatory disorder; presence of atopic dermatitis, senile pruritus, or other skin disease associated with itch; diagnosis of non-CIU chronic urticaria types; and diagnosis of diseases that have urticaria-type symptoms or angioedema. Subjects were also excluded from the study if they had experienced a parasitic or viral infection in the 6 weeks prior to the study screening.

Subjects completed the paper and electronic versions of the UPDD during one 90-min study visit. The study design included a distraction task between administration of the paper and electronic versions of the UPDD in order to minimize the likelihood that subjects would merely recall their earlier UPDD responses. As some of the response options to the UPDD items are based on numbers, it was decided that a distraction task that included numbers would most effectively interfere with recall of UPDD responses. Hence, the distraction task consisted of a series of the popular number puzzle, Sudoku. Subjects were instructed in how to complete the puzzles and were told that, while they were expected to work on the puzzles, they were not required to solve the puzzles. While there is no standard length of delay between repeated administration of a measure in equivalence studies, it is recommended that the delay be long enough to minimize subjects' recall of previous responses yet short enough to minimize changes in subjects' symptoms. Due to the rapidly changing nature of urticaria symptoms, the delay between administration of the UPDD versions was limited to 1 h.

Subjects were randomly assigned to two groups (A and B) sequentially as recruited at each site (Fig. 1). In order to counterbalance order effects, half of the sample (Group A) completed the eUPDD first (T1), followed by an hour of the distraction task, followed by the paper UPDD (T2). The other half of the sample (Group B) completed the paper version first and the electronic version last. The order of the group assignment was alternated by site (e.g., ABAB and BABA). The UPDD consists of two versions, one that is meant to be administered in the morning, and the other that is meant to be administered in the evening. In this study, both the morning and the evening versions were completed at the same time, with the morning version coming first. During the study visit, a trained site staff member

Fig. 1 Subjects were randomly assigned to two groups (A and B) sequentially as recruited at each site. In this way, any possible order effects associated with completing the electronic and paper versions were counterbalanced



conducted informed consent/assent procedures, instructed each subject on the use of both the paper and the electronic version of the UPDD, and moved each subject from task to task in order to ensure adherence to the study protocol.

After completing both versions of the UPDD, subjects were asked to complete a demographic form and three supplementary questions related to any changes in itch severity, number of hives, and size of the largest hive that occurred during the 90-min study visit. Each question had three response options to assess no change, worsening, or improvement in order to determine the stability of each subject's CIU symptomatology (see online supplementary material).

2.1 Electronic Urticaria Patient Daily Diary (eUPDD)

The eUPDD is a daily symptom diary in which subjects rate their symptoms and impacts of urticaria over a 12- or 24-h period. The UPDD was designed to be completed once in the morning and once in the evening, with the items pertaining to the previous 12-h period administered at both time points and the 24-h period items administered at only one time point (see online Supplementary Table 1). Specifically, the morning version contains four items that inquire about itch severity, number of hives, size of the largest hive, and sleep interference, while the evening version contains eight items that inquire about itch severity, number of hives, size of the largest hive, daily activity interference, use of rescue medication, occurrence and treatment of rapid swelling, and calls placed to healthcare professionals.

On the UPDD, patients rate their urticaria symptoms and interference by selecting one of four response options (e.g., none/mild/moderate/severe; none/between 1 and 6 hives/between 7 and 12 hives/greater than 12 hives; none/less than 1.25 cm/between 1.25 and 2.5 cm/greater than 2.5 cm; no interference/mild, little interference/moderate, some interference/substantial, severe interference). Response options for emergency symptom management items are yes/no, number of pills taken, or selecting all that apply from six possible actions (e.g. called my doctor; went to emergency room). Additionally, the UPDD produces one

derived score, the UAS, which is equal to the sum of the 'itch severity' (on a scale where none = 0 to severe = 4) and 'number of hives' scores (none = 0 to 'greater than 12 hives' = 3). The UAS ranges from 0 to 6, with higher scores indicating more symptoms.

The eUPDD was developed for the HTC HD2 Smartphone. The screen size of the HD2 is 3.6 inches long, 2.3 inches wide, and 4.3 inches diagonally, with 480 × 800 WVGA resolution, and the device employs touch-screen technology. While the wording of instructions, items, and response options remained the same between versions, there were some modifications, primarily to formatting and layout. In the paper version, the response options also included the numerical score associated with each option; these were omitted in the electronic version. The items asking about severity of itch and number of hives were presented in two columns, side by side, giving the appearance of being one question in the paper version. In the electronic version, these items were separated and presented as one question per screen. For emphasis, capitalization was used in the electronic version, whereas text was in bold and underlined in the paper version. For purposes of comparisons with an upcoming trial, the drug programmed into the electronic version was diphenhydramine; however, the paper version retained the original drug (hydroxyzine). The software used to program the questionnaires for the eUPDD was tailored using CRF Health's TrialMax platform based on the paper questionnaire. Data collected from the subject entries were encrypted and transferred to a database at CRF Health using standard cellphone technology.

2.2 Sample Size Calculation and Analyses

The minimum sample size required for a valid application of weighted kappa (for equivalency) can be approximated by the formula $2k^2$, where k is the number of categories of classification [12]. With most UPDD items having four categories (i.e., 0–3 response range), the above formula indicates a minimum sample size of 32. Given the variability of urticaria symptoms over short periods of time, the population in this study was oversampled in order to ensure that

data were obtained from a minimum of 32 stable (i.e., no symptom change) patients for the analysis. This target sample size is consistent with the range (20–50 subjects) cited in the meta-analysis of equivalence studies conducted by Gwaltney et al. [13].

As the purpose of this investigation was to assess the equivalence between the paper and electronic versions of the UPDD, only scores for individuals with stable symptoms (i.e., no change) were included in the analyses. Hence, for the ‘itch severity’, ‘number of hives’, and ‘size of largest hive’ UPDD items, analyses were limited to data from those individuals who endorsed the ‘no change’ response option on the corresponding supplemental symptom stability question asking if their symptom had changed since completing the first diary.

Depending upon the response format, the appropriate statistical test was used to assess equivalence. For the six items with four response options each, and one item with six response options, the weighted kappa coefficient was used [14, 15]. Equivalence was evaluated by examining the agreement, as calculated by weighted kappa, between each subject’s item score on the paper UPDD with their score for that item on the electronic UPDD. Interpretation of kappa coefficients was as follows: coefficients less than 0.40 were considered poor, 0.40–0.59 were fair, 0.60–0.74 were good, and those greater than 0.74 were considered excellent [16].

For the two dichotomous items on the UPDD (i.e., presence of rapid swelling and call to doctor/nurse/nurse practitioner), simple kappa coefficients were calculated to determine equivalency between the paper and electronic versions. The UAS was scored based on the sum of the responses to the ‘itch severity’ and ‘number of hives’ items. The covariance and degree of agreement for the morning UAS scores were assessed via intraclass correlation coefficients (ICC) using Shrout–Fleiss reliability fixed set rating [17]; ICCs were also used to test the covariance and degree of agreement between the evening UAS scores. Coons et al. [16] recommend an ICC of at least 0.70 when evaluating equivalence. In addition, mean UAS scores between T1 and T2 were compared using the Wilcoxon signed rank test.

For the item pertaining to rescue medication, equivalency tests were not conducted as the paper and electronic versions of the questionnaire referred to different rescue medications. All data analyses were performed using SAS 9.1 (SAS Institute, Cary, NC, USA).

3 Results

A total of 91 patients participated in the study between October and December 2010; 83 adults and 8 adolescents (see online Supplementary Table 2). The total sample consisted of all subjects from study groups A and B

without any reference to the order of administration of the paper and electronic versions of the UPDD. The mean age of the patients was 43 years (SD 17.2, range 12–74 years), 79 % were female, 76 % identified themselves as White, and the majority of patients (66 %) were employed full- or part-time. The highest education level attained for the majority of patients was some college or a college degree (41 and 30 %, respectively). The mean time since CIU diagnosis was 26 months (SD 82, range 0–336 months).

The number of missing item responses on the paper and electronic versions of the UPDD were very similar and ranged from 0 to 2 %. The response option distributions for both the paper and electronic versions indicated that the sample exhibited the complete range of experiences of symptoms and impacts as assessed by the UPDD.

For the equivalence analyses, the number of patients considered ‘stable’ varied by UPDD item, as each item required stability in a different symptom. With respect to the three main symptoms of urticaria, 74 of 91 patients (81 %) reported no change in itch severity (via the supplemental itch change question) over the course of the 90-min study visit; 67 patients (74 %) reported no change in their number of hives; and 74 patients (81 %) reported no change in the size of their largest hive (as assessed via the respective supplemental questions). A total of 62 of 91 patients (68 %) indicated that neither their itch severity nor their number of hives had changed during the study visit.

Weighted kappa statistics (where weights were applied to deviations from zero) for the three UPDD symptom items are presented in Table 1. All of the weighted kappas were well above the 0.74 recommended threshold, which indicated excellent agreement between the individual item scores obtained on the paper and electronic versions of the UPDD [16]. Perfect agreement between paper and eUPDD in symptom item scores was observed in 83 % (morning itch severity) to 95 % (evening size of largest hive) of cases.

UAS ICCs and Wilcoxon signed rank sum test results are presented in Table 2. The ICCs for the morning and evening UAS were ≥ 0.90 , with Wilcoxon signed rank sum test p values > 0.05 , which indicated good equivalence for the UAS with no significant differences between groups.

Results from the UPDD non-symptom item analyses are presented in Table 3. The values of kappa, some of which were simple (or non-weighted), ranged from 0.82 to 1.00, which again indicated an excellent level of equivalence between the paper and electronic versions of the UPDD.

4 Discussion

The objective of this study was to assess the degree of agreement between results obtained for the Urticaria Patient Daily Diary when administered in paper form

Table 1 Weighted kappas for Urticaria Patient Daily Diary symptom items with four response options

Urticaria patient daily diary item	N	Weighted kappa	Discrepancy between paper and electronic item scores			
			0 (perfect agreement)	±1	±2	±3
Itch severity (morning)						
Stable patients ^a	72	0.8432	60 (83 %)	12 (17 %)	0	0
Itch severity (evening)						
Stable patients ^a	74	0.8908	66 (89 %)	7 (9 %)	1 (1 %)	0
Number of hives (morning)						
Stable patients ^b	65	0.9399	60 (92 %)	5 (8 %)	0	0
Number of hives (evening)						
Stable patients ^b	67	0.9427	62 (93 %)	5 (7 %)	0	0
Size of largest hive (morning)						
Stable patients ^c	60	0.9025	55 (92 %)	5 (8 %)	0	0
Size of largest hive (evening)						
Stable patients ^c	74	0.9427	70 (95 %)	3 (4 %)	1 (1 %)	0

^a Stable patients defined as those who reported on the supplementary question that their itch severity did not change during the study visit

^b Stable patients defined as those who reported on the supplementary question that the number of hives did not change during the study visit

^c Stable patients defined as those who reported on the supplementary question that the size of largest hive did not change during the study visit

Table 2 Intraclass correlation coefficients for Urticaria Patient Daily Diary urticaria activity score

Urticaria Patient Daily Diary	N	Mean (SD) paper version	Mean (SD) electronic version	Difference (paper–electronic)	ICC	Wilcoxon <i>p</i> value
UAS (morning)						
Stable patients ^a	62	3.05 (2.04)	2.94 (1.97)	0.11	0.8977	0.331
UAS (evening)						
Stable patients ^a	62	3.15 (1.98)	3.16 (1.93)	−0.02	0.9509	0.836

ICC ≥ 0.70 is acceptable for group comparisons; Wilcoxon *p* value > 0.05 implies no significant difference between groups

ICC intraclass correlation coefficient, SD standard deviation, UAS Urticaria Activity Score (composite score of number of hives and itch severity; range 0–6)

^a Stable sample defined as those who reported on the supplementary question that no change had occurred during the 90-min study visit with respect to ‘itch severity’ and ‘number of hives’

Table 3 Weighted/simple kappas for supplementary Urticaria Patient Daily Diary items

Urticaria Patient Daily Diary item	N	Weighted or simple kappa	Discrepancy between paper-and-pencil and electronic item scores, N (%)					
			0 (perfect agreement)	±1	±2	±3	±4	±5
Interference with sleep ^a	72	0.8396	62 (86 %)	9 (13 %)	1 (1 %)	0		
Interference with daily activities ^a	74	0.8189	59 (80 %)	15 (20 %)	0	0		
Angioedema ^{a,b}	72	0.9691	71 (99 %)	1 (1 %)				
How did you treat angioedema ^a	17	1.0000	17 (100 %)	0	0	0	0	0
Call doctor, nurse, or nurse practitioner ^a	73	0.8817	72 (99 %)	1 (1 %)				

^a Data are for stable patients; defined as those who reported on the supplementary question that their itch severity did not change during the study visit

^b Only simple kappa can be calculated due to dichotomous response options

versus when administered electronically. Formal equivalency testing was warranted, according to FDA PRO guidelines, because the eUPDD contained moderate modifications to the appearance of the original instrument (due to changes in layout and the screen size of the electronic hand-held device used).

Using specifically designed questions capturing no change, worsening or improvement in three key symptoms, subsamples of patients with stable urticaria symptomatology were identified and the data derived from these subsamples were used to evaluate equivalence. The results of the equivalency testing showed that CIU patients' responses to all UPDD items (symptom, non-symptom, and UAS) were consistent across paper versus electronic administration methods.

A meta-analysis of equivalence studies between paper- and screen-based electronic administration of PRO measures revealed that 94 % of correlations (most of which were ICCs) were at least 0.75, with the average correlation equal to 0.90 [13]. The findings from the current study (ICCs ≥ 0.90 and kappas from 0.82 to 1.00) are in line with those from other equivalence studies.

The study has some limitations. The time between administrations was necessarily short given the potentially rapid change in the number and size of hives and related itch that could have occurred. While a distraction task was included to minimize the likelihood of direct recall, recall may have been a contributing component of the observed agreement between versions. Additionally, it should be noted that the assessment of change between administrations in the number and size of hives was based on patient self-report as opposed to objective measures. Thus, patients may not have accurately identified themselves as being stable versus having improved or worsened with regard to these two signs. Further, this study included only patients with CIU, and additional research would be needed to confirm whether the findings are generalizable to other types of urticaria.

5 Conclusions

As the use of electronically administered questionnaires continues to increase, it is important for researchers to ensure that electronic versions of PRO measures yield results that are equivalent to those obtained with the validated paper versions of the measures. The UPDD, originally developed in paper format, has been successfully migrated to an electronic version. In this study, a diverse sample of patients with CIU completed both versions of the UPDD during one 90-min study visit. Item-level analyses provided strong evidence to support the measurement

equivalence of the electronic and paper versions of the UPDD.

Acknowledgments The authors thank Jennifer Tucker and Kathleen Beusterien for their assistance in drafting the manuscript. The study was funded by Genentech. Jim Zazzali is employed by Genentech. Emuella M. Flood and Jennifer Devlen were contracted by Genentech to perform the study

Authors' contributions JZ conceived the study and participated in its design. EF participated in the design, coordination, and analysis of the study. JD participated in the statistical analysis of the study. All authors read and approved the final manuscript. EF is the guarantor for the overall content.

References

1. Ferrer M, Kaplan AP. Progress and challenges in the understanding of chronic urticaria. *Allergy Asthma Clin Immunol*. 2007;3(1):31–5.
2. Di Lorenzo G, D'Alcamo A, Rizzo M, et al. Leukotriene receptor antagonists in monotherapy or in combination with antihistamines in the treatment of chronic urticaria: a systematic review. *J Asthma Allergy*. 2009;2:9–16.
3. Zuberier T, Bindslev-Jensen C, Canonica W, et al. EAACI/GA²LEN/EDF guideline: definition, classification and diagnosis of urticaria. *Allergy*. 2006;61:316–20.
4. Mlynec A, Zalewska-Janowska A, Martus P, et al. How to assess disease activity in patients with chronic urticaria? *Allergy*. 2008;63:777–80.
5. Wan KS. Efficacy of leukotriene receptor antagonist with an anti-H1 receptor antagonist for treatment of chronic idiopathic urticaria. *J Dermatol Treat*. 2009;20(4):194–7.
6. Kaplan AP, Joseph K, Maykut RJ, et al. Treatment of chronic autoimmune urticaria with omalizumab. *J Allergy Clin Immunol*. 2008;122(3):569–73.
7. Serhat Inaloz H, Ozturk S, Akcali C, et al. Low-dose and short-term cyclosporine treatment in patients with chronic idiopathic urticaria: a clinical and immunological evaluation. *J Dermatol*. 2008;35(5):276–82.
8. Mathias SD, Dreskin SC, Kaplan A, et al. Development of a daily diary for patients with chronic idiopathic urticaria. *Ann Allergy Asthma Immunol*. 2010;105:142–8.
9. US Department of Health and Human Services Food and Drug Administration FDA Guidance for Industry. Patient-reported outcome measures: use in medical product development to support labeling claims; 2009. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>.
10. Mathias SD, Tschosik EA, Zazzali JL. Adaptation and validation of the Urticaria Patient Daily Diary for adolescents. *Allergy Asthma Proc*. 2012;33:186–90.
11. Mathias SD, Crosby RD, Zazzali JL, et al. Evaluating the minimally important difference of the urticaria activity score and other measures of disease activity in patients with chronic idiopathic urticaria. *Ann Allergy Asthma Immunol*. 2012;108:20–4.
12. Cicchetti DV. Testing the normal approximation and minimal sample size requirements of weighted kappa when the number of categories is large. *App Psychol Meas*. 1981;5(1):101–4.
13. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome

- measures: a meta-analytic review. *Value Health*. 2008;11(2):322–33.
14. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70:213–20.
 15. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas*. 1973;33:613–9.
 16. Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO good research practices task force report. *Value Health*. 2009;12(4):419–29.
 17. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–8.