

Original article

The reliability, validity and responsiveness of the International Restless Legs Syndrome Study Group rating scale and subscales in a clinical-trial setting

Linda Abetz^{a,*}, Robert Arbuckle^a, Richard P. Allen^b, Diego Garcia-Borreguero^c,
Wayne Hening^d, Arthur S. Walters^e, Elena Mavraki^f, Jeffrey M. Kirsch^f

^a Mapi Values, Adelphi Mill, Grimshaw Lane, Bollington, Cheshire SK10 5JB, UK

^b Johns Hopkins Bayview Medical Center, Baltimore, MD, USA

^c Fundación Jiménez Díaz, Madrid, Spain

^d UMDNJ-RW Johnson Medical School, New Brunswick, NJ, USA

^e The New Jersey Neuroscience Institute at JFK Medical Center, Seton Hall University School of Graduate Medical Education, Edison, NJ, USA

^f GlaxoSmithKline, Greenford, UK

Received 24 June 2005; received in revised form 27 December 2005; accepted 29 December 2005

Abstract

Patients and methods: To assess the reliability, validity, and responsiveness of the International Restless Legs Syndrome Study Group's rating scale (the International Restless Legs Scale (IRLS)) (V2.0), using pooled data from two matching, placebo-controlled studies of ropinirole for treating Restless Legs Syndrome (RLS).

Results: Pooled patient samples comprised 550 patients in the baseline (validation) sample and 439 patients in the week 12 longitudinal (responsiveness) sample. Factor analysis revealed acceptability of the IRLS total score (accounting for 40% of the variance) and that nine of the 10 IRLS items could also be assigned to two distinct subscales, the symptoms or symptoms impact subscales. The IRLS total score, symptoms and symptoms impact subscales had acceptable construct validity, internal consistency reliability ($\alpha=0.81$, 0.80, and 0.76, respectively), and concurrent validity ($r=-0.68$, -0.52 , -0.70 , respectively, with the Restless Legs Syndrome Quality of Life questionnaire (RLSQoL) overall life impact score). IRLS scores differed significantly between different levels of sleep problems and Clinical Global Impression (CGI) of health status ($P<0.0001$), indicating known groups and clinical validity, respectively. Changes in scores differed significantly among CGI 'global improvement' levels ($P<0.0001$), providing evidence of responsiveness.

Conclusions: The IRLS total score, symptoms, and symptoms impact subscales are reliable, valid, and responsive in a clinical trial setting.

© 2006 Published by Elsevier B.V.

Keywords: IRLS; Restless legs syndrome; Reliability; Validity; Responsiveness; Psychometric analysis

1. Introduction

Restless legs syndrome (RLS) is a neurological disorder characterized by an urgent need to move the limbs (most often the legs) when the patient sits or lies down, usually accompanied by paresthesias (unpleasant sensations, such as 'creeping', 'crawling', 'tingling', 'pulling', or 'pain').

Moving the limbs brings rapid, if variable, relief from the symptoms, but the relief tends to last only as long as the movement continues [1].

The prevalence of RLS increases with age, and the rate in women is about twice that for men [2]. The overall prevalence of RLS appears to vary quite widely, from 2.5 to 15%, depending on the population surveyed [3,4]. There are a number of differential diagnoses, such as leg cramps, paresthesias due to peripheral neuropathy, and arthritic or muscular pain [5]. There are also three major causes of secondary RLS: renal failure, pregnancy, and iron deficiency anemia. Primary RLS has a tendency to run in families. Recent genetic linkage and association studies

* Corresponding author.

E-mail address: linda.abetz@mapivalues.com (L. Abetz).

have identified possible areas for a susceptibility gene on chromosomes 12q, 9p, and 14q [6–8]. Each of these susceptibility loci occurs in some RLS families, but not the majority.

RLS becomes worse at night, and clinically significant RLS is usually associated with disruptions to circadian pattern and sleep impairment on a regular basis, leading to fatigue, poor concentration, anxiety, or depression and compromised quality of life [9–13]. It is important, therefore, that measures developed to assess the severity of RLS take into account not only the symptoms themselves, but also the impact of RLS on sleep, mood, and daily functioning. Two disease-specific, clinician-administered measures of RLS symptom severity have been developed and validated: the Johns Hopkins Restless Legs Severity Scale (JHRLSS) [14], and the International Restless Legs Scale (IRLS), developed by the International Restless Legs Syndrome Study Group [1,15]. The JHRLSS was designed as a limited clinical guide based on time of symptom onset. The IRLS, on the other hand, is a more comprehensive measure, consisting of 10 items that address a range of RLS symptoms and their impact on patients' mood and daily life.

Although the original IRLS (Version 1.0) has already been validated [15], it is vital to ensure that the questionnaire also performs well psychometrically when used in different patient groups and as the instrument is further refined through general use. The psychometric properties of the IRLS total score were, therefore, assessed in the patient samples of two recently completed phase-III, multicenter, randomized, double-blind, placebo-controlled studies assessing the efficacy and tolerability of ropinirole, a dopamine agonist, for the treatment of adults with moderate-to-severe RLS: TREAT RLS 1 (Therapy with Ropinirole; Efficacy And Tolerability in RLS 1 [16]), and TREAT RLS 2 [17]. The primary endpoint in both studies was change in IRLS total score.

The findings of two separate psychometric analyses of these studies confirmed the validity of the IRLS total score as the primary measure of overall RLS severity and yielded subscales that were similar to those noted previously [15,18]. The aim of the present study, therefore, was to provide an assessment of the reliability, validity and responsiveness of the IRLS total score and the two potential subscale scores, in a trial patient sample based on the TREAT RLS 1 and 2 studies. The data from both studies were pooled in order to increase the statistical power of the analyses.

2. Methods

2.1. Patient samples

The patient samples from TREAT RLS 1 and 2 were pooled for the present psychometric analysis. Patients were eligible for inclusion in each study if they were at least 18 years of age and had moderate-to-severe RLS (had a

baseline IRLS total score of >15 and either had experienced at least 15 nights with symptoms of RLS in the previous month or, if receiving treatment, had symptoms of this frequency before treatment). Patients were excluded from the study if they had any other movement or primary sleep disorder, if they required daytime treatment for RLS, if they were experiencing augmentation or end-of-dose rebound, or if they had secondary RLS. Patients were also excluded if they had a history of alcohol or drug abuse, previous intolerance to dopamine agonists, or were suffering from other clinically relevant conditions affecting assessments.

All patients gave written, informed consent before entering the studies, done according to the principles of the 1996 amendment of the declaration of Helsinki and approved by local ethics committees.

2.2. Clinical trial study design

As matching study designs were used for both studies, it was considered appropriate to pool the data for this analysis. The studies were conducted in a double-blind, randomized, placebo-controlled manner. Patients were recruited from hospitals, sleep centers and neurology clinics in 10 European countries in TREAT RLS 1 (Austria, Belgium, France, Germany, Italy, The Netherlands, Norway, Spain, Sweden and the UK) and in six countries around the world in TREAT RLS 2 (Australia, Canada, Germany, Norway, the UK and the USA). Patients receiving treatment for RLS or treatment known to affect RLS or sleep, or to cause drowsiness, entered a washout phase of either seven consecutive nights or five half-lives of the drug, whichever was the greater. Patients were randomized in a 1:1 ratio to receive once-daily treatment with either ropinirole or placebo for 12 weeks. Ropinirole was initiated at a dose of 0.25 mg/day and titrated upwards during weeks 1–7, either until they were judged to have reached their optimal dose or until they reached the maximum dose of 4.0 mg/day. During the titration period, a maximum of two dose reductions was permitted in the case of adverse events, and doses could then be increased again if the adverse events improved. No further dose changes were permitted after week 7.

The primary endpoint in both studies was change in the IRLS total score, as published previously [16,17]. Secondary endpoints included Clinical Global Impression (CGI) 'global improvement' and 'severity of illness' scores, change in the Restless Legs Syndrome Quality of Life questionnaire (RLSQoL) score, and the medical outcomes study sleep problems index II (MOS sleep scale) score.

2.3. Outcome measures used in psychometric analysis

2.3.1. IRLS

The IRLS was developed and validated by the International Restless Legs Syndrome Study Group [1,15,18].

The scale consists of 10 questions concerning each patient's symptoms and the impact of those symptoms on daily activities and mood. The version used in TREAT RLS 1 and 2 (IRLS V2.0) was adapted to have a 1-week recall, and minor grammatical modifications were made to ease the translation into various linguistic versions required for the clinical studies. Each question contains answers that score from 0 to 4 points, with 0 representing the absence of a problem and 4 representing a very severe problem. The IRLS was completed at baseline, Day 2 and weeks 1–8 and 12 of the treatment phase, or at time of withdrawal for patients who discontinued the studies prematurely. Change in the IRLS total score from baseline to week 12 was the primary endpoint in both studies. In addition to the IRLS total score, the two potential IRLS subscales measuring 'symptoms' and 'symptoms impact', as identified from the factor analysis (see Section 3), were also subjected to psychometric analysis.

2.3.2. CGI

The CGI consists of three modules, the CGI 'global improvement' (CGI-I), the CGI 'severity of illness' (CGI-S), and the CGI 'efficacy index', and has been in use for nearly three decades [19]. In the present two studies, only the first two modules were used as outcome measures, although the CGI 'efficacy index' was used by the investigators to guide titration of the study medication. The CGI-I and CGI-S modules were assessed by the investigator, based on all information available at the time of rating. Both modules were rated on a scale of 0–7. For the CGI-I, 0 refers to patients who were not assessed, 1 indicates 'very much improved', and 7 indicates 'very much worse'. For the CGI-S, 0 refers to patients who were not assessed, 1 indicates 'normal, not at all ill' patients, and 7 refers to patients who were 'among the most extremely ill' patients. Changes in the proportions of patients with scores of 'much improved' or 'very much improved' were identified as two key secondary endpoints. Both the CGI-I and CGI-S were assessed by the investigators at Day 2 and weeks 1–8 and 12 of the treatment phase, or at the time of withdrawal in patients who discontinued the study prematurely. The CGI-S was also assessed at baseline.

2.3.3. MOS sleep scale

The MOS sleep scale is a self-administered scale measuring specific aspects of sleep (i.e. problems with sleep disturbance [initiation and maintenance], adequacy, somnolence, quantity, respiratory impairments, and snoring) and has been found to be reliable and valid in the general US population [20,21]. It was designed for use in patients who may have varying co-morbidities, and hence is appropriate for a medically diverse patient population. In addition to the sleep problems index II summary score, only the subscales of sleep disturbance, sleep adequacy, somnolence, and sleep quantity were calculated, as they were considered the most appropriate to assess sleep impairment in RLS patients. The frequency with which each problem has been experienced during the previous 4

weeks is scored on a six-point scale ranging from 'none of the time' to 'all of the time', except for sleep quantity, which is measured in hours. Patients were asked to complete the MOS sleep scale at baseline, and at weeks 8 and 12 of the treatment phase, or at the time of withdrawal for patients who discontinued the study prematurely. The psychometric properties of the MOS sleep scale have also been assessed within each of these two clinical trial samples and found to be satisfactory.

2.3.4. RLSQoL

The RLSQoL is a validated questionnaire consisting of 18 items, of which 13 are scored on a five-point scale, the remainder being recorded as either a numerical value or a dichotomous response [10]. Ten of the items contribute to a single summary score, the overall life impact score, whereas the remaining eight items concern employment (one question), sexual interest (two questions), and work (five questions) and are summarized individually. Thus, the RLSQoL addresses the impact of RLS on daily activities, social well-being, work, sex life, and the ability to concentrate and make decisions. Higher scores on the RLSQoL overall life impact score indicate a better quality of life. Patients were asked to complete the RLSQoL at baseline and at weeks 8 and 12 of the treatment phase, or at time of withdrawal for patients who discontinued the studies prematurely. The psychometric properties of the RLSQoL have been assessed previously, and within each of these two clinical trial samples, and found to be satisfactory [22,23].

2.4. Psychometric analysis

2.4.1. Study populations

Analysis was performed on patients from the intention-to-treat (ITT) sample; that is, all randomized patients who received at least one dose of study medication and who had at least one post-baseline efficacy measurement. Patients from the ITT sample who had an evaluable IRLS score (i.e. at least nine non-missing items) were included in the IRLS baseline validation sample, which was used for all psychometric analyses of the IRLS total score and subscales, except responsiveness. Patients included in the IRLS baseline validation sample who also had an evaluable IRLS score at a post-baseline visit were included in the IRLS longitudinal validation sample, which was used for analysis of the responsiveness to change over time of the IRLS total score and subscales. All tests were performed on the total pooled sample, blinded to treatment status.

2.4.2. Psychometric validation of the IRLS total score and subscales

In order to assess the a priori structure of the IRLS, exploratory factor analysis was carried out using oblique principal component analysis (PCA), with the number of factors left free. The resulting item-scale structure was

Table 1
Psychometric analyses: purpose of tests

Property	Purpose
Item convergent validity	To assess an item's correlation with its hypothesized subscale score (satisfied if correlation achieved is ≥ 0.40) [29]
Item discriminant validity	To assess whether an item considered in isolation has a higher correlation with its hypothesized scale than with the other scale in the questionnaire [29]
Internal consistency reliability	To evaluate the extent to which individual items of the instrument are consistent with each other and reflect an underlying scheme or construct (satisfied if Cronbach's α coefficient ≥ 0.70 is achieved) [30,31]
Concurrent validity	To assess correlations between the IRLS and IRLS subscales and other, validated, measures (in this case, the RLSQoL and the MOS sleep scale) [32]. Correlations ≥ 0.40 among similar scales are considered sufficient evidence of concurrent validity
Known groups validity	To confirm that the IRLS total score and subscales can distinguish between known groups [32]; in this case, IRLS scores were compared between mild, moderate, and severe sleep problems as defined by taking tertile scores for the sleep problems index II of the MOS sleep scale
Clinical validity	To confirm the IRLS total score and subscale scores can discriminate between patients who differ in their clinical status [32]. Assessed by examining the correlation with CGI-S, a measure of overall health status (a correlation of ≥ 0.40 was considered satisfactory), and by comparing IRLS total and subscale scores among CGI-S subgroups
Responsiveness	To assess responsiveness of a measure to change over time [32]. Correlations between the change in IRLS total score and subscale scores and CGI-I scores were assessed ($r \geq 0.40$ considered satisfactory). In addition, changes in IRLS scores were compared among CGI-I levels. The effect size (ES) was used as a measure of change: In the range of 0.2, small; 0.5, moderate and 0.8, large [25,26]

assessed for the following psychometric properties: item convergent validity, item discriminant validity, internal consistency reliability, concurrent validity, known groups validity, clinical validity, and responsiveness. Table 1 briefly summarizes the purpose and criterion for each of these tests.

Concurrent validity of the IRLS total score and subscale scores was assessed by examining correlations with the RLSQoL overall life impact score and the sleep problems index II of the MOS sleep scale, both of which measure related health concepts such as those assessed by the IRLS. When interpreting results, correlations are described as negligible if < 0.20 , small if ≥ 0.20 and < 0.40 , moderate if ≥ 0.40 and < 0.70 , and large if ≥ 0.70 . The same ranges were used in the interpretation of correlations evaluated in the assessment of clinical validity and responsiveness.

The known groups validity of the IRLS was assessed by describing and comparing IRLS total scores and subscale scores at baseline among groups of mild, moderate, and severe sleep problems, as defined by taking tertile scores for the sleep problems index II of the MOS sleep scale. Tertile scores were used because no pre-defined clinical cut-offs are available for the MOS sleep scale. Taking tertile scores involves dividing a normally distributed sample into three groups of as close as possible to 33% of patients in each group. Scores of 0–41, 42–56, and 57–100 were considered mild, moderate, and severe, respectively. The hypothesis was that patients with more severe sleep problems would have worse IRLS total scores and subscale scores.

The clinical validity of the IRLS was assessed in two ways. First, correlations between baseline CGI-S scores and IRLS total and subscale scores were assessed. The IRLS total and subscale scores were expected to be moderately related to the clinician rating of health status (CGI-S); therefore, correlations of ≥ 0.40 were considered sufficient

to confirm validity. Second, IRLS total scores and subscale scores at baseline were compared among subgroups with different levels of overall health status, defined by stratifying the CGI-S scores into three groups: scores 1–2 (normal, not at all ill, or borderline ill); scores 3–5 (mild, moderately or markedly ill); and scores 6–7 (severely ill, or among the most extremely ill patients). Patients with a CGI-S score of 0 ('not assessed') were excluded from this analysis. The hypothesis was that patients with worse clinician-rated overall health status (CGI-S) would also have worse IRLS total scores and subscale scores.

Responsiveness to change over time was assessed in the longitudinal validation sample by evaluating correlations between change in IRLS total and subscale scores (calculated by subtracting baseline from post-baseline assessments) and CGI-I scores, at each post-baseline assessment.

In addition, changes in IRLS total and subscale scores from baseline to week 12 were compared among the seven CGI-I groups (from 1 to 7, representing 'very much improved' through to 'very much worse'). Patients with a CGI-I score of 0 ('not assessed') were excluded from the analysis. The effect size (ES) was used as a measure of the change in IRLS scores within each CGI-I group. ESs were calculated by dividing the change in mean IRLS total scores and subscale scores (from baseline to a subsequent time point) by the SD of mean scores at baseline. The ES has been recommended in the literature as an appropriate benchmark for evaluating the magnitude and meaning of change in health status measures [24].

Cohen defined effect sizes of 0.2, 0.5, and 0.8 as small, moderate, and large, respectively [25]. We adopted Guyatt et al.'s guidance that size effects can be described as small, moderate or large when results are in the range of these parameters [26].

2.4.3. Statistics

No adjustments for multiplicity were performed. The Type 1 error was 0.05, and all hypothesis tests were two-sided. Where correlations were evaluated, Spearman correlation coefficients were calculated. Differences in IRLS total scores and subscale scores among pairs of groups were assessed using Mann–Whitney–Wilcoxon tests. Kruskal–Wallis tests were used in the comparison of more than two groups.

3. Results

3.1. Patients

The pooled patient samples from the two studies comprised 550 patients in the baseline validation sample and 439 patients in the week 12 longitudinal (responsiveness) sample. Baseline patient characteristics are shown in Table 2. The majority of patients were categorized as moderately, markedly, or severely ill. The mean age of patients was 55.3 ± 11.1 years, the mean age of onset of symptoms was 36.4 ± 17.1 years, and approximately two-thirds of patients were women. This reflects the population seeking treatment for RLS.

Table 2
Pooled patient characteristics: baseline validation sample ($N=550$)

Characteristic	
Age, years	
Mean (SD)	55.3 ± 11.1
Range	28.0–79.0
Sex, % (N)	
Men	38.5 (212)
Women	61.5 (338)
Work status, % (N)	
Full-time employment	39.6 (218)
Part-time employment	14.0 (77)
Volunteer/unpaid work	2.2 (12)
Unemployed due to RLS	0.9 (5)
Unemployed due to other (non-RLS)	1.8 (10)
symptoms	
Unemployed for other reasons	6.4 (35)
Retired	25.6 (141)
Homemaker	9.5 (52)
CGI 'Severity of Illness', % (N)	
Missing data	0.2 (1)
Normal, not at all ill	1.3 (7)
Borderline ill	1.3 (7)
Mildly ill	8.2 (45)
Moderately ill	35.5 (195)
Markedly ill	31.8 (175)
Severely ill	19.6 (108)
Among the most extremely ill patients	2.2 (12)
Age of onset of symptoms, years	
Mean (SD)	36.4 ± 17.1

3.2. Missing data

For patients included in the IRLS baseline validation sample, no missing data were observed for any of the items on the IRLS.

3.3. Factor analysis

The cumulative variance of the first factor (unrotated) was 0.40, equaling the cumulative variance standard, indicating that it is appropriate for an IRLS total score to be calculated. The factor analysis also suggested two potential subscales (one containing five items pertaining to symptoms and one containing five items pertaining to impact of symptoms on life). Rotated factor coefficients are presented in Table 3. Rotated factor coefficients ranged from 0.61 to 0.78 for the five items in the 'Symptoms' subscale, indicating that all items loaded satisfactorily with their own subscale or factor. On the other hand, rotated factor coefficients for the 'Symptoms Impact' subscale (range 0.22–0.81) indicated that one of the five items did not load with its own factor (rotated factor coefficient = 0.22). This finding suggests that the item in question (IRLS item 3, 'Overall, how much relief of RLS arm/leg discomfort do you get from moving around?') is measuring a distinct concept, unrelated to the other items.

Of the nine remaining items, all except one loaded only with their own factor. The exception (IRLS item 4, 'How severe was your sleep disturbance due to RLS?') had rotated factor coefficients of 0.49 with the 'Symptoms' subscale, and 0.55 with the 'Symptoms Impact' subscale, suggesting that it might be best placed in the latter subscale. However, previous factor analysis of the IRLS suggested that this item should be placed in the 'Symptoms' subscale [18], and this was supported by advice from the expert RLS clinicians who are co-authors of this study, confirming the face validity of this solution. Additional analyses to assess whether the discriminative power of the subscales differed depending on which subscale item 4 was placed in were conducted. Results of these additional analyses (available from the authors on request) were very similar regardless of where item 4 was placed. Therefore, for the purposes of the present psychometric validation report, analyses have been presented for a 'Symptoms' subscale comprising six items (items 1, 2, 4, 6, 7 and 8), and a 'Symptoms Impact' subscale comprising three items (items 5, 9 and 10).

3.4. Psychometric analysis

The results of the psychometric evaluations demonstrated acceptable reliability and validity of the IRLS total score and subscales (Tables 4 and 5; Figs. 1–3).

3.4.1. Item convergent validity

All except two items were found to satisfy the standard for item convergent validity (≥ 0.40) for the IRLS total

Table 3
Rotated factor coefficients for the IRLS subscale items, baseline validation sample ($N=550$)

IRLS item ^a		Factor 1 symptoms	Factor 2 symptoms impact
Symptoms subscale			
6	How severe was your RLS as a whole?	0.78499	0.35471
1	Overall how would you rate the RLS discomfort in your legs or arms?	0.76071	0.31520
2	Overall how would you rate the need to move around because of your RLS symptoms?	0.70441	0.08537
8	When you had RLS symptoms, how severe were they on average?	0.63245	0.26180
7	How often did you get RLS symptoms?	0.60790	−0.02173
4 ^b	How severe was your sleep disturbance due to your RLS symptoms?	0.48976	0.54722
Symptoms impact subscale			
9	Overall how severe was the impact of your RLS symptoms on your ability to carry out daily affairs?	0.13909	0.81171
5	How severe was your tiredness or sleepiness during the day due to your RLS symptoms?	0.09064	0.79774
10	How severe was your mood disturbance due to your RLS symptoms?	0.13760	0.77947
Single item			
3	Overall how much relief of your RLS arm/leg discomfort did you get from moving around?	0.09403	0.21768

^a Patients were asked to recall over the past week for each item.

^b Despite a slightly greater loading on Factor 2 (Symptoms Impact), item 4 was included in Factor 1 (Symptoms) for the purposes of this evaluation—see Section 3.3 for rationale.

score (Table 4). The exceptions were items 3 ('Overall how much relief of your RLS arm/leg discomfort did you get from moving around?') and 7 ('How often did you get RLS symptoms?'), which had correlations of 0.16 and 0.31, respectively. For the IRLS symptoms subscale, the only item that did not satisfy the standard for item convergent validity was item 7, which had a correlation of 0.37. For the IRLS symptoms impact subscale, all items surpassed the standard criterion for item convergent validity.

3.4.2. Item discriminant validity

All items in the IRLS subscales met the standard for item discriminant validity (Table 4). That is, all items had a higher item-subscale correlation with their own IRLS subscale rather than the other IRLS subscale, providing further evidence that item 4 belongs in the symptom subscale.

3.4.3. Internal consistency reliability

Cronbach's α coefficients for the IRLS total score, symptoms, and symptoms impact subscales scores were 0.81, 0.80, and 0.76, respectively, all exceeding the minimum reliability standard of ≥ 0.70 for internal consistency reliability (Table 4).

3.4.4. Concurrent validity

Concurrent validity was confirmed by the statistically significant correlations between the IRLS total score, both subscale scores, and the RLSQoL overall life impact score ($P < 0.0001$ in all three cases) (Table 5). The correlations were moderate for the IRLS total score and the symptoms subscale ($r = -0.68$ and -0.52 , respectively), and large for the symptoms impact subscale ($r = -0.70$). The higher correlation for the symptoms impact subscale compared to the symptoms subscale was expected, as the symptoms

Table 4
Validity and reliability of the IRLS and IRLS subscales, baseline validation sample ($N=550$)

	% Success	Coefficient range
Item convergent validity (% items with item-scale correlation ≥ 0.40)		
Total score	88.9	0.16–0.69
Symptoms subscale	83.3	0.37–0.75
Symptoms impact subscale	100	0.55–0.65
Item discriminant validity (% items correlated more highly with their own subscale than with the other subscale)		
Symptoms subscale	100	–
Symptoms impact subscale	100	–
Internal consistency reliability (% of scales with Cronbach's α coefficient ≥ 0.70)		
Total score	–	0.81
Symptoms subscale	–	0.80
Symptoms impact subscale	–	0.76

Table 5
Concurrent validity of the IRLS total score and subscale scores

	Total score	Symptoms subscale score	Symptoms impact subscale score
RLSQoL overall impact score			
<i>N</i>	547	547	547
Correlation	−0.68	−0.52	−0.70
<i>P</i> value	<0.0001	<0.0001	<0.0001
MOS sleep scale, sleep problems index II			
<i>N</i>	550	550	550
Correlation ^a	0.61	0.49	0.59
<i>P</i> value	<0.0001	<0.0001	<0.0001

^a Spearman correlation coefficient.

impact subscale is more likely to reflect the impact of RLS on quality of life, rather than the symptoms subscale.

Correlations between the IRLS total score and subscale scores, and the sleep problems index II of the MOS sleep scale were all moderate and all statistically significant ($P < 0.0001$ in each case) (Table 5), providing further confirmation of the concurrent validity of the IRLS total score and subscales, and indicating that sleep problems are closely related to both RLS symptoms and their impact on a patient’s life.

3.4.5. Known groups validity

Known groups validity of the IRLS total score, the IRLS symptoms subscale, and the IRLS symptoms impact subscale was demonstrated by the significant differences in scores according to severity of sleep problems (MOS sleep scale, sleep problems index II) ($P < 0.0001$ in each case) (Fig. 1). That is, the IRLS total score and subscales are able to distinguish between groups of patients with different levels of sleep problems.

3.4.6. Clinical validity

Clinical validity was supported by the statistically significant correlations between the IRLS total score,

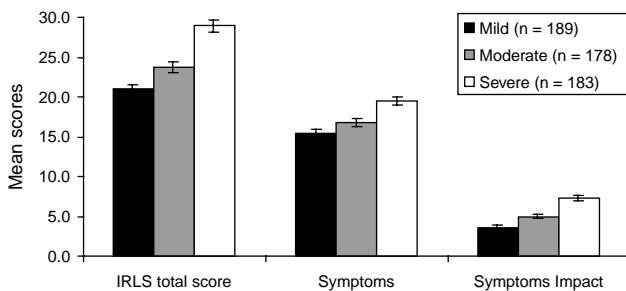


Fig. 1. Known groups validity: comparison of IRLS total score and subscale scores according to severity of sleep problems. $P < 0.0001$ for mild versus moderate, mild versus severe, and moderate versus severe for IRLS total score, IRLS symptoms subscale score, and IRLS symptoms impact subscale score (Mann–Whitney–Wilcoxon test). Mild, moderate, and severe groups based on upper, middle, and lower tertile scores from the sleep problems index II of the MOS sleep scale.

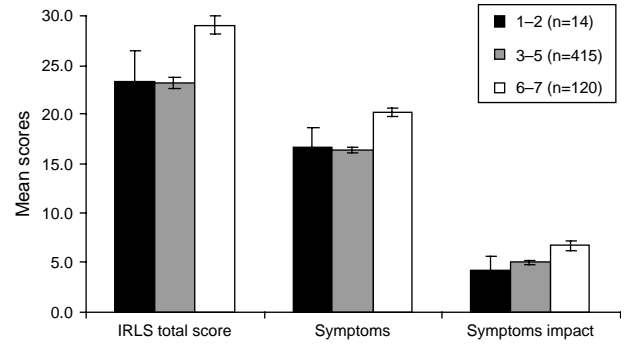


Fig. 2. Clinical validity: comparison of IRLS total and subscale scores among CGI-S groups. $P < 0.0001$ for 3–5 versus 6–7, $P < 0.01$ for 1–2 versus 6–7, and $P = NS$ for 1–2 versus 3–5 for IRLS total score, IRLS symptoms subscale score, and IRLS symptoms impact subscale score (Mann–Whitney–Wilcoxon test). CGI-S group 1–2 includes ‘normal, not at all ill’ and ‘borderline ill patients’; group 3–5 included ‘mild’, ‘moderate’, and ‘markedly ill’ patients; and group 6–7 includes patients who were ‘severely ill’ or ‘among the most extremely ill patients’.

subscale scores and the CGI-S. The moderate correlation between the IRLS total score and the CGI-S was identical to that between the IRLS symptoms subscale score and the CGI-S ($r = 0.57$, $P < 0.0001$ in each case). For the IRLS symptoms impact subscale score, the correlation with CGI-S fell just below the 0.40 threshold but remained statistically significant ($r = 0.39$, $P < 0.0001$). When scores were compared between CGI-S subgroups, the CGI-S group 6–7 had significantly worse IRLS total and subscale scores than both the CGI-S groups 3–5 and 1–2 (Fig. 2). For CGI-S group 1–2 versus group 3–5, the differences were not significant. However, the sample sizes in the 1–2 group were small in this study due to the inclusion criteria.

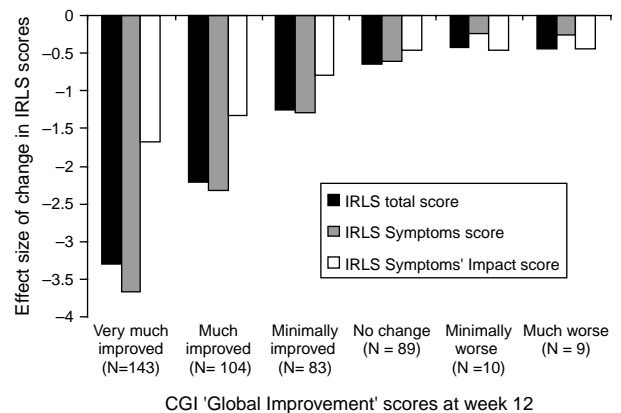


Fig. 3. Responsiveness as determined by change in IRLS total and subscale scores from baseline to week 12: effect sizes are compared among CGI-I groups. The single patient with a CGI-I score of ‘very much worse’ is not included, as effect sizes cannot be calculated for one person. $P < 0.0001$ for comparisons among groups for the IRLS total score, symptoms and symptoms impact scores (Kruskal–Wallis test). Effect sizes were considered in the range of 0.2, small; 0.5, moderate; and 0.8, large [25,26].

3.4.7. Responsiveness

The responsiveness of the IRLS total score and subscales to change over time was confirmed by moderate-to-large correlations between the change in the IRLS total score and subscale scores and CGI-I scores at each post-baseline time point. The correlations of the CGI-I with the IRLS total score ranged from $r=0.71$ to 0.74 , correlations with the symptoms subscale ranged from 0.71 to 0.75 , and correlations with the symptoms impact subscale ranged from 0.44 to 0.53 .

Furthermore, changes in IRLS total scores and both IRLS subscale scores (from baseline to week 12) differed between CGI-I levels (1–7) at a statistically significant level ($P<0.0001$, ES range = -0.23 to -3.67) (Fig. 3). Improvements in the IRLS total score and subscale scores were larger for the ‘more improved’ CGI-I groups compared with the ‘less improved’ and ‘no change’ groups, indicating the responsiveness of the IRLS total score and subscales. For the IRLS total score and both IRLS subscales, statistically significant differences were found for comparisons of the ‘very much improved’ group versus all other CGI-I groups ($P<0.01$ in all cases), for ‘much improved’ versus all other groups ($P<0.05$ in all cases), and for ‘minimally improved’ versus ‘no change’ ($P<0.01$ in all cases). Differences between CGI-I ‘no change’ and any of the CGI-I groups which had become worse were not statistically significant for the IRLS total score or for either of the subscales. However, patient numbers in the ‘minimally worse’, ‘much worse’, and ‘very much worse’ CGI-I groups were very small (10, 9, and 1, respectively).

For the IRLS total score and both subscales, there was a step-wise increase in effect sizes for the ‘no change’, and ‘minimally improved’, ‘much improved’, and ‘very much improved’ groups, indicating greater improvements in IRLS scores for the more improved CGI-I groups.

4. Discussion

Based on the results of this psychometric evaluation, the IRLS total score and both the IRLS symptoms subscale score and the IRLS symptoms impact subscale score were found to be reliable, valid, and responsive to change. These findings support the use of the IRLS total score and subscale scores with patients suffering from RLS in a clinical trial setting.

Factor analysis revealed that all items except one met the standard criterion for correlation with their designated subscale (rotated factor coefficient ≥ 0.40). The exception was item 3 (‘Overall, how much relief of RLS arm/leg discomfort do you get from moving around?’), which failed to meet the criterion for either subscale, suggesting that it might be considered a distinct concept. Discussions with clinical experts involved in this validation study suggested the usefulness of item 3 as a diagnostic criterion but not necessarily as an indicator of symptom severity. This may

explain why the item did not correlate particularly well with items known to indicate RLS severity. Item 4 (‘How severe was your sleep disturbance due to your RLS symptoms?’) was the only item to load >0.40 on both subscale scores, which is perhaps not surprising as sleep disturbance is an important symptom of RLS and can also have a major impact on patients’ daily life. Kushida and colleagues demonstrated that the impact of RLS on daily activities is primarily due to sleep disturbance [27]. As previous factor analysis of the IRLS suggested that item 4 be included in the symptoms subscale [18], and there was support for this from the RLS expert clinicians involved in this validation study, it was placed in the symptoms subscale for the present analysis. However, although not included in this article in detail, psychometric analyses were also performed with item 4 included in the alternative symptoms impact subscale, and the findings were similar to those obtained when it was included in the symptoms subscale.

In terms of item convergent validity, only items 3 and 7 (‘How often did you get RLS symptoms?’) did not meet the 0.40 criterion for correlation with the IRLS total score (0.16 and 0.31 , respectively). The lower item-scale correlation for item 7 may be largely due to the inclusion criteria for the study, which required subjects to have had 15 or more RLS episodes in the last month. Consequently, all but five patients gave the top three responses for this item, whereas the responses for the other items were more normally distributed. This reduced variability may have resulted in a lower correlation for item 7 with the other IRLS items. Nevertheless, since both items provide important information to the clinician, they were retained in the total score, although item 3 was omitted from both subscales.

All items in both IRLS subscales met the standard for item discriminant validity, correlating more highly with their own subscale than with the other subscale. These results provide further evidence that it is appropriate to calculate the subscale scores.

The IRLS total score and both subscale scores all exceeded the standard for internal consistency reliability, providing evidence of their reliability. The standard of 0.40 for concurrent validity was also exceeded in correlations between the IRLS total or subscale scores and both the RLSQoL overall impact score and the sleep problems index II of the MOS sleep scale. The levels of correlation confirmed the concurrent validity of the IRLS and subscales with these previously validated measures, indicating that the measures are assessing overlapping, but not identical, concepts. The finding that the symptoms impact subscale correlated more strongly with the RLSQoL than the symptoms subscale (-0.70 versus -0.52) is consistent with the item composition of the respective subscales.

The IRLS total score and both subscale scores were able to distinguish between differing levels of sleep problems at baseline, as determined by the sleep problems index II of the MOS sleep scale. That is, for patients with more disturbed sleep, IRLS total and subscale scores were worse, thereby

confirming their known groups validity. Their clinical validity was assessed by looking at correlations with baseline CGI-S scores. Both the IRLS total score and the symptoms subscale score exceeded the correlation standard of 0.40 (0.57 in each case), whereas the symptoms impact subscale correlation fell just below the standard at 0.39 but remained statistically significant ($P < 0.0001$). The higher correlation of CGI-S with the symptoms subscale compared with the symptoms impact subscale suggests that clinicians are more likely to be assessing the symptomatology of RLS rather than its impact on the individual when determining the overall health status of patients with RLS.

When differences in IRLS scores were compared among patients with different degrees of overall health status according to CGI-S categorization, patients in the '6–7' group had significantly worse IRLS total scores, symptoms scores and symptoms impact scores than both the '3–5' and '1–2' groups. The latter could not be distinguished statistically, but sample sizes in the '1–2' group were very small due to the inclusion criteria. Further research is warranted comparing a larger sample size of RLS patients with a broad range of overall health status. With this caveat, it was considered that the IRLS total score and both subscale scores were clinically valid at an acceptable level.

Responsiveness to change over time was demonstrated by significant, moderate-to-large correlations between changes in IRLS total and subscale scores and CGI-I scores. In addition, the changes in IRLS total scores and subscale scores from baseline to week 12 differed among the seven CGI-I scores at a statistically significant level. Effect sizes for patients rated by their clinicians as improved were large, effect sizes for patients rated as stable were moderate to large, and effect sizes for patients rated as having become worse were negligible to moderate (where sample sizes were sufficiently large to be examined). The patient numbers in the 'worsened' group were very small and any findings should be viewed with caution.

Improvements in the 'no change' group would suggest the presence of a Hawthorne effect, i.e. a response shift due simply to experience with the questionnaire and potentially due to the positive psychosocial effects of participating in a clinical trial (regardless of treatment). These improvements may also be due to the difference in perspective between the two measures. The IRLS assesses a patient's perspective, whereas the CGI is the clinician's overall impression. Of note, the greater improvements in terms of effect size observed in the symptom scale versus the symptom-impact scale may suggest that symptoms are more likely to be immediately responsive to treatment than symptom impact. This has certainly been shown in other chronic disease areas where patients can immediately see, understand and report symptoms relief, but may be less able to recognize that they can now do things that they could not do previously [28].

4.1. Study limitations

In terms of limitations, the clinical studies were designed to assess the efficacy and safety of ropinirole in the treatment of RLS and not designed with the validation of the IRLS as the primary aim. For this reason, the study did not include any unaffected control subjects or patients with mild RLS; all patients included had moderate or severe RLS. This was confirmed by the small number of patients who were rated by their clinician as being 'normal, not at all ill', 'borderline ill' or 'mildly ill' at baseline (1.27, 1.27, and 8.18%, respectively, on the CGI-S). Therefore, the findings of the study described here cannot be generalized to patients with mild RLS or no RLS symptoms. Given that all IRLS items attribute symptoms to RLS, it would be inappropriate to give the scale to those who do not have a diagnosis of RLS. However, it should be kept in mind that the IRLS has been previously validated in a study that included patients with RLS at all levels of severity [15]; the main purpose of this study was to further demonstrate the validity of the IRLS in a clinical trial sample.

In addition, the mean age of the participants was 53 ± 11 years, indicating that younger RLS patients were not well represented. This most likely reflects the increasing prevalence of RLS with age and the inclusion criteria of the trials in terms of severity [2]. Nevertheless, further validation of the IRLS in a sample in which younger patients are better represented is warranted.

Finally, there were very few patients who experienced aggravation of their RLS severity (either in terms of clinician rating or patient report) over the course of the study. Further study of the responsiveness of the IRLS to aggravation in RLS severity is therefore also warranted.

5. Conclusions

The findings of this study support the conclusion that the IRLS total score, the primary overall measure of RLS severity, is valid, reliable, and responsive to improvements in RLS severity, in patients suffering from RLS, in a clinical trial setting. However, the IRLS symptoms and symptoms impact subscales are also reliable, valid, and responsive in this setting and may be used to assess the impact of treatments on both symptoms and the impact these symptoms have on patients with RLS. Further research may be warranted to assess the responsiveness of the IRLS to worsening of RLS severity.

Acknowledgements

Both the clinical trials and the psychometric analysis were supported by GlaxoSmithKline Research and Development.

References

- [1] Allen RP, Picchietti D, Hening WA, et al. Restless legs syndrome diagnosis and epidemiology workshop at the national institutes of health; international restless legs syndrome study group. Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. *Sleep Med* 2003;4:101–19.
- [2] Allen RP, Walters AS, Montplaisir J, et al. Restless Legs Syndrome prevalence and impact: REST general population study. *Arch Intern Med* 2005;165:1286–92.
- [3] Allen RP, Earley CJ. Restless legs syndrome: a review of clinical and pathophysiologic features. *J Clin Neurophysiol* 2001;18:128–47.
- [4] Earley CP. Restless legs syndrome. *N Engl J Med* 2003;348:2103–9.
- [5] Hening WA. Subjective and objective criteria in the diagnosis of the restless legs syndrome. *Sleep Med* 2004;5:285–92.
- [6] Desautels A, Turecki G, Montplaisir J, et al. Identification of a major susceptibility locus for restless legs syndrome on chromosome 12q. *Am J Hum Genet* 2001;69:1266–70.
- [7] Chen S, Ondo WG, Rao S, et al. Genomewide linkage scan identifies a novel susceptibility locus for restless legs syndrome on chromosome 9p. *Am J Hum Genet* 2004;74:876–85.
- [8] Bonati MT, Ferini-Strambi L, Aridon P, et al. Autosomal dominant restless legs syndrome maps on chromosome 14q. *Brain* 2003;126(Pt 6):1485–92.
- [9] Winkelmann J, Prager M, Lieb R, et al. ‘Anxietas Tibiarum’: depression and anxiety disorders in patients with restless legs syndrome. *J Neurol* 2005;252:67–71.
- [10] Abetz L, Allen R, Follet A, et al. Evaluating the quality of life of patients with restless legs syndrome. *Clin Ther* 2004;26:925–35.
- [11] Saletu B, Gruber G, Saletu M, et al. Sleep laboratory studies in restless legs syndrome patients as compared with normals and acute effects of ropinirole. 1. Findings on objective and subjective sleep and awakening quality. *Neuropsychobiology* 2000;41:181–9.
- [12] Rothdach AJ, Trenkwalder C, Haberstock J, et al. Prevalence and risk factors of RLS in an elderly population: the MEMO study. Memory and morbidity in Augsburg elderly. *Neurology* 2000;54:1064–8.
- [13] Allen R, Kirsch J, Abetz L, Washburn T. The impact of restless legs syndrome (RLS) on sleep and cognitive functioning. *Eur J Neurol* 2002;9(Suppl. 2):50 [SC334].
- [14] Allen RP, Earley CJ. Validation of the Johns Hopkins Restless Legs Severity Scale (JHRLSS). *Sleep Med* 2001;2:239–42.
- [15] Walters AS, LeBrocq C, Dhar A, et al. Validation of the International Restless Legs Syndrome Study Group Rating Scale for restless legs syndrome. *Sleep Med* 2003;4:121–32.
- [16] Trenkwalder C, Garcia-Borreguero D, Montagna P, et al. Ropinirole in the treatment of restless legs syndrome: results from the TREAT RLS 1 study, a 12-week, randomized, placebo controlled study in 10 European countries. *J Neurol Neurosurg Psychiatry* 2004;75:92–7.
- [17] Walters AS, Ondo WG, Dreykluft T, et al. Ropinirole is effective in the treatment of restless legs syndrome: TREAT RLS 2: a 12-week, double-blind, randomized, parallel-group, placebo-controlled study. *Mov Disord* 2004;(19):1414–23.
- [18] Allen RP, Kushida CA, Atkinson MJ. RLS QoL Consortium. Factor analysis of the International Restless Legs Syndrome Study Group’s scale for restless legs severity. *Sleep Med* 2003;4:133–5.
- [19] Guy W, editor. Clinical global impression (CGI). ECDEU assessment manual for psychopharmacology. Rockville, MD: US Department of Health and Human Services, Public Health Service, Alcohol Drug Abuse and Mental Health Administration, NIMH Psychopharmacology Research Branch; 1976. p. 218–22.
- [20] Hays RD, Stewart AL. Sleep measures. In: Stewart AL, Ware Jr JE, editors. Measuring functioning and well-being. The medical outcomes study approach. Durham: Duke University Press; 1992. p. 235–59.
- [21] Hays RD, Martin SA, Sesti AM, Spritzer KL. Psychometric properties of the Medical Outcomes Study sleep measure. *Sleep Med* 2005;6:41–4.
- [22] Abetz L, Arbuckle R, Allen R, et al. Validating the Restless Legs Syndrome Quality of Life questionnaire (RLSQoL) in a trial patient population. *Value Health* 2004;7(6):793.
- [23] Abetz L, Vallow S, Kirsch J, et al. Validation of the restless legs syndrome quality of life questionnaire. *Value Health* 2005;8:157–67.
- [24] Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;27(3 Suppl.):178–89.
- [25] Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Lawrence Erlbaum Associates; 1988.
- [26] Guyatt GH, Osaba D, Wu AW, et al. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002;77:371–83.
- [27] Kushida CA, Allen RP, Atkinson MJ. Modeling the causal relationships between symptoms associated with restless legs syndrome and the patient-reported impact of RLS. *Sleep Med* 2004;5:485–8.
- [28] Juniper EF, Svensson K, O’Byrne PM, et al. Asthma quality of life during 1 year of treatment with budesonide with or without formoterol. *Eur Respir J* 1999;14:1038–43.
- [29] Campbell DT, Fiske DW. Convergent and discriminant validation by the Multi-trait Multi-method Matrix. *Psychol Bull* 1959;56:81–105.
- [30] Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
- [31] Nunnally JC. Psychometric theory. 2nd ed. New York: McGraw-Hill; 1978.
- [32] Hays RD, Anderson R, Revicki DA. Assessing reliability and validity of measurement in clinical trials. In: Staquet M, Hays R, Fayers PM, editors. Quality of life assessment in clinical trials. Methods and practice. Oxford: Oxford University Press; 1998.