



## Original Article

Validation of the self-administered version of the international Restless Legs Syndrome study group severity rating scale – The sIRLS<sup>1</sup>

Denise Sharon<sup>a,\*</sup>, Richard P. Allen<sup>b</sup>, Pablo Martinez-Martin<sup>c</sup>, Arthur S. Walters<sup>d</sup>, Luigi Ferini Strambi<sup>e</sup>, Birgit Högl<sup>f</sup>, Lynn Marie Trotti<sup>g</sup>, Mark Buchfuhrer<sup>h</sup>, John Swieca<sup>i</sup>, Richard K. Bogan<sup>j</sup>, Rochelle Zak<sup>k</sup>, Jennifer G. Hensley<sup>l</sup>, Laurel A. Schaefer<sup>m</sup>, S. Marelli<sup>n</sup>, Marco Zucconi<sup>n</sup>, Ambra Stefani<sup>o</sup>, Evi Holzknecht<sup>o</sup>, Victoria Olvera<sup>p</sup>, Hailey Meaklim<sup>i</sup>, Irena Laska<sup>i</sup>, Philip M. Becker<sup>q</sup>, International RLS Study Group<sup>r</sup>

<sup>a</sup> Tulane University School of Medicine, Clinical Director Advanced Sleep Center Metairie Louisiana, PVMC Sleep Disorders Center, Claremont, CA, USA

<sup>b</sup> School of Medicine, Johns Hopkins University, Baltimore, MD, USA

<sup>c</sup> National Center for Epidemiology, Carlos III Institute of Health, Madrid, Spain

<sup>d</sup> Division of Sleep Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA

<sup>e</sup> Department of Neurology, OSR-Turro, Department of Neuroscience, Director Sleep Disorders Center, University Vita-Salute San Raffaele, Milan, Italy

<sup>f</sup> Sleep Disorders Center, Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria

<sup>g</sup> Emory University, Atlanta, GA, USA

<sup>h</sup> Stanford University School of Medicine, Palo Alto, CA, USA

<sup>i</sup> Melbourne Sleep Disorders Center, East Melbourne, Australia

<sup>j</sup> University of South Carolina School of Medicine, Columbia, SC, Sleep Med, Inc, USA

<sup>k</sup> University of California San Francisco Sleep Disorders Center, San Francisco, CA, USA

<sup>l</sup> University of Texas Austin, School of Nursing, Austin, TX, USA

<sup>m</sup> Vanderbilt University, Nashville, TN, USA

<sup>n</sup> Department of Neuroscience, Sleep Disorders Center, University Vita-Salute San Raffaele, Milan, Italy

<sup>o</sup> Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria

<sup>p</sup> Emory Healthcare, Atlanta, GA, USA

<sup>q</sup> Southwestern University School of Medicine, Medical Director Sleep Medicine Associates of Texas, Dallas, TX, USA

<sup>r</sup> The International Restless Legs Syndrome Study Group, Rochester, MN, USA

## ARTICLE INFO

## Article history:

Received 9 July 2018

Received in revised form

6 October 2018

Accepted 16 October 2018

Available online 29 October 2018

## Keywords:

Restless Legs Syndrome

RLS symptoms

Severity scale

Self-administered

Patient completed

Validity/reliability

## ABSTRACT

**Introduction:** The International Restless Legs Study Group (IRLSSG) has developed the IRLS (International Restless Legs Syndrome Severity Scale) and validated it as a clinician/researcher administered scale to be used when both patient and examiner are present. The IRLSSG recognized the need for a self-completing scale that can be used economically in clinical practice and in large population-based studies. In this study the validity and the reliability of the IRLS as a self-administered scale (sIRLS) is assessed.

**Methods:** Established RLS patients were recruited by eight centers in four countries and consented to participate in this study. The validity of the sIRLS was assessed by patients completing the sIRLS before a clinician administered the IRLS. The reliability of the sIRLS was assessed by patients completing the sIRLS again, two weeks after the first one, provided no change had occurred.

**Results:** Overall, 173 patients were recruited and 164 of them were included in the analyses. The sIRLS showed satisfactory scaling assumptions and no relevant floor or ceiling effect. One factor explained 61.3% of the variance. Cronbach's alpha was 0.93 and the item homogeneity index was 0.59. Intraclass correlation coefficient between the sIRLS and the IRLS was 0.94. The sIRLS standard error of measurement was 3.61 (½ SD at baseline = 4.11). The results mostly overlapped those of the IRLS analyzed in parallel.

**Discussion:** The sIRLS is a reliable, valid and precise instrument that showed tight association with the IRLS. These findings support the use of the sIRLS for self-evaluation of RLS severity. The responses

\* Corresponding author.

E-mail address: [denisesharon23@gmail.com](mailto:denisesharon23@gmail.com) (D. Sharon).

<sup>1</sup> The self-reporting IRLS tool (sIRLS) is protected by international copyright, with all rights reserved to IRLSSG Foundation. Do not use without permission. For information on, or permission to use the scale, please contact its official licensor and distributor Mapi Research Trust: <https://eprovide.mapi-trust.org/>.

obtained on the sIRLS and the IRLS scale varied slightly. Therefore, we recommend that either the sIRLS or the IRLS scale be used as the only scale for serial measures over time.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

The International Restless Legs Severity Rating Scale (IRLS) was developed by the International Restless Legs Syndrome Study Group (IRLSSG) and validated in 2003 [1]. Since then, it has served as the major endpoint in most academic studies where Restless Legs Syndrome (RLS) severity is measured and most industry sponsored trials where the United States Food and Drug Administration (FDA) and as well as its European counterpart approval is sought for the use of medications in RLS [2–6]. The IRLS is a widely used measure for evaluation of RLS severity. A Web-of-Science search as of May 2018, lists 812 publications where the scale is either employed or discussed. The IRLS shows robust clinimetric properties under both therapeutic and non-therapeutic conditions [1,7–9] and compares favorably to other severity rating scales for RLS [10–12].

The IRLS was validated only under conditions where an examiner had to be available to answer any questions the patient may have about the items in the scale [1]. With the advent of large questionnaire-based and epidemiology studies in RLS, it has become apparent that there is a need for a rating scale that can be answered by the patient in the clinic, by mail or on-line without assistance from a clinician. The availability of validated measures to assess treatment efficacy is crucial to performance-based medicine.

The current study is undertaken to determine if the IRLS with modified instructions (sIRLS) can be administered without the presence of an examiner and to compare the robustness of the scale when administered with (IRLS) and without the presence of an examiner (sIRLS).

## 2. Methods

### 2.1. Design

This is an open, cross-sectional, observational, multicenter, international study in a cohort of English speaking patients diagnosed with Restless Legs Syndrome by expert clinicians.

### 2.2. Patients

Patients were recruited from eight participating centers, five in different parts of the US (ASW, LMT, MB, RB, RZ), one in Italy (LFS), one in Austria (BH) and one in Australia (JS). Each center obtained approval from its institutional review board or from a centralized review board. All participating investigators have previously initiated and participated in clinical studies involving RLS patients, are well known in their community for their RLS expertise and have good control of the English language. All patients were previously diagnosed with RLS in the respective center. Centers were instructed to recruit patients with a wide range of RLS symptoms. There was no limit on how many patients a center could recruit between 2015 and 2017. A total of 217 patients were considered for this study and 173 agreed to participate.

### 2.3. Inclusion and exclusion criteria

Inclusion criteria were: (1) an established diagnosis of RLS based on the revised IRLSSG criteria [13]; (2) stable medications for past 2

weeks and willingness to continue on the same regimen without a foreseen adverse effect; (3) age 18 and older; (4) able to read, understand and answer written questions in English as ascertained by the center investigators; and (5) able to provide informed consent.

Exclusion criteria included not matching any of the inclusion criteria and retraction of the informed consent.

### 2.4. Procedures

Written informed consent was obtained from each patient and the protocol was explained to the subject with an emphasis on the need to maintain a similar lifestyle with no changes in medication for the two-week duration of the study. On day one, RLS subjects who consented to participate in the study completed the sIRLS and received stamped envelopes containing the sIRLS and the PGIC with the center clinician's address. On the same day following the completion of the initial sIRLS by the subject, a clinician familiar with the scale and blind to the sIRLS results administered the IRLS. The patient was then instructed to wait two weeks (one week with clinician's approval if there was an urgent need for medication adjustment or change) before completing and mailing the questionnaires included in the envelope, with re-iteration of the need to maintain the same lifestyle and medication regimen for the two-week duration of the study. If medication adjustments were needed at the time of enrollment, participants were asked to maintain current medications for one week and then complete the questionnaires prior to changing medications, ie, they had only one week between questionnaire administrations. If medication changes were necessary in less than one week, the subject was removed from the study. If the changes were requested at the one-week time point or later, the subject was requested to fill out the forms just prior to changing medications. The subject was instructed to complete the questionnaires at approximately the same time as on day one (morning, afternoon, or evening). Once completed, the questionnaires were returned to the center via mail, fax, or e-mail. Questionnaires returned within 7–21 days were included in the analyses. Phone, text, or email reminders were provided to guarantee the completion of the questionnaires and return of the envelope.

### 2.5. Assessments

Socio-demographic and historical RLS data from each patient in the study were recorded. In addition, the following assessments, all in English were applied: sIRLS, IRLS and Patient Global Impression of Change (PGIC).

1. The sIRLS (copyrights by IRLSSG)<sup>1</sup> is a self-evaluation composed of 10 items that replicates the model of the IRLS (see below), with adaptation of the wording for self-application by the subjects (Appendix 1). The wording was agreed upon by the authors after considering comments and suggestions from members of the RLS Foundation who are RLS patients and did not participate in the study. The scale includes explanations and instructions for the user, and the time frame it assesses is “the past 7 days”. The items score from 0 (no symptoms) to 4 (very severe symptoms), and a total sIRLS score (0–40) is obtained by the sum of the item scores.

2. The IRLS is a rater-based measure of RLS severity [11], widely used and considered the primary endpoint for most clinical trials. The IRLS includes 10 items scored from 0 to 4, and the total score runs from 0 (no symptoms) to 40 (the most severe symptoms). The time frame is “the past week”. The IRLS has been thoroughly validated, showing excellent clinimetric properties, and has been recommended for evaluation of RLS severity by the *ad hoc* Movement Disorder Society Task Force [11].
3. The modified Patient Global Impression of Change (PGIC) is a tool that includes a question on change (1 = Condition got worse; 2 = No change; and 3 to 7 reflecting increasing levels of improvement) and a numerical visual scale from 0 (much better) to 10 (much worse) [14]. In our study the PGIC was used to monitor the stability of the RLS condition since the first application of the sIRLS until the second application scheduled for the test-retest reliability analysis. Thus, it was essential to have a stand-alone “no change” item.

## 2.6. Data analysis

Encoded data were locally entered in a database and submitted to the National Center of Epidemiology (Madrid, Spain) to be merged in an anonymized central database designed and maintained following international and local norms for safety, confidentiality, and ethical requirements.

In addition to descriptive statistics for socio-demographic data and assessment scores (mean, standard deviation, median, range, percentage, etc.), the following analyses were carried out:

- 1 Review of data quality to determine the proportion of full computable scores. Five percent was considered the limit for missing data [15].
- 2 According to the Shapiro-Francia test, the sIRLS and IRLS total scores and age-related variables showed a normal distribution and, therefore parametric tests were used for these data. Floor and ceiling effects were determined as acceptability parameters (limit,  $\leq 15\%$ ) [16].
- 3 Scaling assumptions parameters were explored by rough equivalence of means, standard deviations, and standard error of the means. Corrected item-total correlation that was deemed satisfactory if it was  $>0.30$  [17].
- 4 The dimensionality of the scale was examined by exploratory factor analysis, principal component factor. Adequacies of the variables and sample were checked with Bartlett's and Kaiser-Meyer-Olkin's tests.
- 5 Internal consistency was analyzed through the Cronbach's alpha coefficient (an estimate of reliability reflecting all possible correlations between the items of a scale; minimal acceptable value, 0.70), inter-item correlation ( $\geq 0.20$ ), item homogeneity coefficient (mean of the inter-item correlations;  $>0.30$ ), and the aforementioned item-total correlation [15,18,19]. The abovementioned attributes were simultaneously explored for the sIRLS and the IRLS for allowing the reader to compare these properties between both scales.
- 6 Test-retest reliability was carried out in those patients who declared “no change” between the first and second application according to the PGIC (question and visual analog scale). Weighted kappa with quadratic weights (STATA v. 15) for items and ICC (1-way, random effect) for the total scores were used for this purpose.
- 7 The sIRLS precision was determined by the standard error of measurement (SEM) on the test-retest reliability index [20].
- 8 The statistical significance of the difference between sIRLS and IRLS scores was determined through Wilcoxon test (for

the items) and the student *t*-test for the total scores. The agreement between the sIRLS (patient-based) and the original IRLS (examiner-based) was analyzed using weighted kappa with prerecorded quadratic weights for items, intra-class correlation coefficient (ICC) (2-way, random effect) and Bland–Altman plot for the total scores.

- 9 The concurrent validity between the original IRLS and the sIRLS was established with the Pearson's product-moment correlation coefficient for the total scores and with the Spearman rank correlation coefficients for the individual items. Pearson's correlation coefficient was also used to explore the association of the sIRLS and IRLS with age, age at onset, age at start of treatment, and RLS duration.
- 10 For determining the frequency of the RLS symptoms included in the sIRLS and IRLS, scores = 0 were considered “symptom absent” and scores  $\geq 1$ , “symptom present”. A comparison between the respective proportions was carried out using the McNemar's exact test.

## 3. Results

### 3.1. Description of the sample

One hundred seventy-three patients were included in the study. Nine of them did not complete the sIRLS and, therefore, they were not considered for the analysis. One patient had two missing values in the scale (items 9 and 10; 0.1% of 1640 sIRLS item scores) that were imputed by the individual's mean of the other items. Therefore, the sIRLS analysis was carried out on data from 164 patients, 61.6% women. Descriptive statistics of the sample are shown in Table 1 and the sIRLS and IRLS scores are shown in the Table 2.

### 3.2. Data quality and acceptability

Nine patients lacked IRLS data. Therefore, the sample for the analyses referred to this scale is  $n = 155$ . All items of both scales covered the 0–4 theoretical range up to a possible total score of 40. The total scores were 0–38 for the sIRLS and 0–36 for the IRLS. No relevant skewness ( $-0.11$ ) nor floor or ceiling effect was present in the sIRLS total score (3.1% and 0.6%, respectively) nor in the IRLS (5.2% and 2.6%), but some items in both scales showed floor or ceiling effect higher than the standard 15% (Table 3).

### 3.3. Scaling assumptions

Mean, standard deviation, and standard error of the mean) were quite similar among the sIRLS item scores. The maximum difference observed for these statistics were 1.35 (means), 0.30 (SD), and 0.04 (standard error). Item-to-total corrected correlation ranged from 0.46 (item 3) to 0.88 (item 6), clearly higher than the criterion 0.30. Similar results were found with the IRLS both for score distribution and item-total correlations (Table 3).

**Table 1**  
Descriptive statistics of the sample.

	Mean	SD	Range
Age	59.48	13.19	24–87
Age at onset	38.32	17.47	5–77
Age at diagnosis	49.92	14.58	13–83
Age at start of treatment	52.17	13.43	10–83
Restless Legs Syndrome duration	20.64	17.14	0–76

All values in years.

**Table 2**  
Scores of the sIRLS and IRLS.

Items	sIRLS (n = 164)			IRLS (n = 155)		
	Median	Mean	SD	Median	Mean	SD
1. Discomfort	2	2.05	1.06	2	2.03	1.10
2. Need to move	2	2.09	1.10	2	2.10	1.45
3. Relief	2	1.79	0.98	2	1.76	1.01
4. Sleep disturbance	2	1.85	1.23	2	1.74	1.24
5. During the day	2	1.65	1.20	1	1.47	1.20
6. RLS on the whole	2	1.98	0.99	2	2.01	1.06
7. How often	3	2.73	1.24	3	2.68	1.26
8. How severe	2	2.08	0.94	2	2.01	0.99
9. Daily activities	1	1.38	1.17	1	1.23	1.20
10. Mood disturbance	1	1.39	1.19	1	1.33	1.25
Total score	20	18.99	8.83	19	18.49	9.33

sIRLS: Self Completed Restless Legs Syndrome Rating Scale.

IRLS: Restless Legs Syndrome Rating Scale.

SD: Standard deviation.

### 3.4. Dimensionality

The exploratory factor analysis of the sIRLS identified only one factor explaining 61.3% of the variance (Bartlett test of sphericity,  $p < 0.001$ ; Kaiser-Meyer-Olkin measure, 0.93). This factor, interpreted as “perceived severity of the IRLS symptoms” included all items of the sIRLS (loadings from 0.52, item 3, to 0.91, item 6; mean: 0.79). A similar result was obtained with the IRLS: one factor explaining 66.2% of the variance (Bartlett test of sphericity,  $p < 0.001$ ; Kaiser-Meyer-Olkin measure, 0.92).

### 3.5. Internal consistency

sIRLS Cronbach's alpha was 0.93 and the item homogeneity index was 0.59. Inter-item correlations ranged from 0.29 (item 3 with item 10) to 0.88 (item 1 with item 6). The respective values for the IRLS were 0.33 (item 3 with item 4) and 0.87 (item 1 with item 6).

### 3.6. Stability and precision of the sIRLS

Test-retest reliability of the sIRLS was explored in 63 patients who remained stable (median: 14 days; interquartile range: 14–15) according to the PGIC (question and visual analog scale) and resulted satisfactory for the total score (ICC = 0.82). For items the average weighted kappa value was 0.65, ranging from 0.38 (item 3) to 0.80 (item 6). The SEM based on the test-retest ICC was 3.48 (½ SD at baseline, 4.11).

**Table 3**  
Data of acceptability and Item-total correlation.

Items	sIRLS (n = 164)				IRLS (n = 155)			
	Skew.	Floor	Ceiling	I-T corr.	Skew.	Floor	Ceiling	I-T corr.
1. Discomfort	0.19	4.9	10.4	0.86	0.08	7.7	11.0	0.88
2. Need to move	-0.01	6.7	10.4	0.82	-0.01	8.4	12.9	0.79
3. Relief	0.27	6.1	3.7	0.46	0.34	7.1	4.5	0.45
4. Sleep disturbance	0.21	15.2	12.8	0.74	0.17	20.0	9.7	0.81
5. During the day	0.22	20.1	6.7	0.72	0.42	24.5	5.2	0.72
6. RLS on the whole	-0.04	6.7	5.5	0.88	0.14	6.5	9.7	0.89
7. How often	-0.45	3.7	39.6	0.70	0.44	5.2	38.1	0.70
8. How severe	-0.02	3.1	4.9	0.78	0.01	5.2	5.8	0.80
9. Daily activities	0.40	29.9	4.3	0.80	0.50	39.4	3.2	0.81
10. Mood disturbance	0.28	31.1	2.4	0.67	0.44	36.1	4.5	0.72
Total score	-0.11	3.1	0.6	-	-0.04	5.2	2.6	-

sIRLS: Self Completed Restless Legs Syndrome Rating Scale.

IRLS: Restless Legs Syndrome Rating Scale.

Skew.: Skewness.

Floor and Ceiling: Floor and Ceiling effect (in %).

I-T corr.: Corrected Item-Total correlation.

### 3.7. Comparison and agreement between sIRLS and IRLS

Significant differences were found for items 4 ( $p = 0.01$ ), 5 ( $p = 0.001$ ), 9 ( $p = 0.002$ ), and total score ( $p = 0.016$ ), but differences higher than  $\pm 1$  point for each item were observed in  $<95\%$  of the cases for all items, with full agreement ranging from 68.4% (item 10) to 83.9% of patients (item 6) (Table 4). Agreement of scores from sIRLS and IRLS according to weighted kappa and ICC tests was substantial or better ( $\geq 0.80$ ) (Table 5). After verifying that item scores had a normal distribution (Shapiro-Francia test), the Bland-Altman analysis was applied to each item. According to this procedure, the number of cases (and %) falling out of the respective 95% CI ranged from 2 (1.3%, for item 6) to 7 (4.5%, for items 2 and 9) and the mean difference between scales was  $0.61 \pm 3.12$  (95% confidence interval:  $-5.50$  to  $6.73$ ). The correlation of the difference in score between scales and the mean score of the two scales was  $-0.13$  (a value near zero indicates concordance). There were 11 values (12 cases; 7.7% of the sample) out of the 95% CI: seven lower and five higher than the mentioned limits. The plot showed that the sIRLS tended towards scoring slightly higher than the IRLS, and that the differences between the scores of both scales, were located across all severity levels, without a systematic distribution (Fig. 1).

### 3.8. Concurrent validity and association with other variables

The correlation between individual sIRLS and IRLS components and total scores reached values from 0.80 to 0.94 (Table 5). Both scales showed negligible correlations ( $<0.10$ ) with age, age at RLS onset, age at diagnosis, age at start of treatment, and duration of the disorder. Although both total scores were mildly higher in women, the differences by gender did not reach statistical significance.

### 3.9. Frequency of RLS symptoms captured by sIRLS and IRLS

The frequency of RLS individual symptom items according to both scales appears in Table 6. As a whole, there was a trend to find lower rates with the IRLS than with the sIRLS, but the difference was significant only for item 9.

### 3.10. Cross-cultural issues

Comparisons of results from participating centers in English speaking countries (engl) and participating centers in Europe (eu) are presented as Supplementary Material Tables 1–7. The samples were significantly different by number: 97 vs. 67, age: mean 62.4 (11.1) vs. 55.2 (14.7), and RLS duration in years: mean 26.2 (18.1) vs

**Table 4**  
Individual item scores difference between sIRLS and IRLS.

Difference in score	Number of patients									
	-4	-3	-2	-1	0	1	2	3	4	
1. Discomfort	0	0	1	16	120	16	1	1	0	
2. Need to move	0	1	4	14	117	17	1	1	0	
3. Relief	0	0	3	13	118	18	2	1	0	
4. Sleep disturbance	0	0	3	10	112	27	2	0	1	
5. During the day	0	0	2	13	101	35	2	2	0	
6. RLS on the whole	0	0	1	13	130	10	1	0	0	
7. How often	0	0	0	10	127	13	5	0	0	
8. How severe	0	0	0	13	120	19	2	1	0	
9. Daily activities	0	0	1	11	110	27	5	1	0	
10. Mood disturbance	0	0	2	18	106	25	3	1	0	

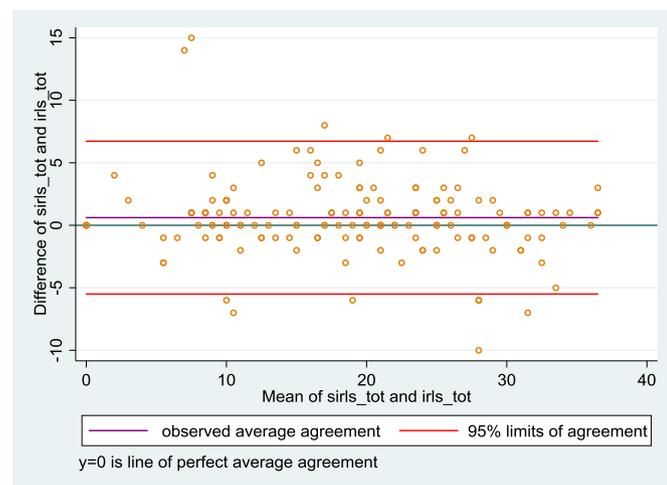
sIRLS: Self Completed Restless Legs Syndrome Rating Scale.  
IRLS: Restless Legs Syndrome Rating Scale.

**Table 5**  
Agreement, and correlation between the sIRLS and IRLS.

Items	Agreement sIRLS – IRLS	Correlation sIRLS – IRLS
	$k_w$	Spearman
1. Discomfort	0.86	0.87
2. Need to move	0.82	0.83
3. Relief	0.80	0.80
4. Sleep disturbance	0.85	0.86
5. During the day	0.82	0.82
6. RLS on the whole	0.91	0.92
7. How often	0.91	0.91
8. How severe	0.83	0.84
9. Daily activities	0.84	0.84
10. Mood disturbance	0.84	0.84
Total score	0.94 <sup>a</sup>	0.94 <sup>b</sup>

$k_w$ : Weighted kappa coefficient.  
Spearman: Spearman correlation coefficient.  
<sup>a</sup> Intraclass correlation coefficient.  
<sup>b</sup> Pearson's correlation coefficient.

20.6 (17.1). Gender distribution was similar in both samples. When scales were completed in the office, the mean scores on the sIRLS and IRLS were very similar, but significant differences were noted when English speaking patients were compared to European patients: sIRLS – 16.9 (8.3) vs. 22.0 (8.8) and IRLS – 16.0 (8.6) vs. 22.6



Differences between sIRLS and IRLS total scores:  $0.61 \pm 3.12$  (95%CI: -5.50, 6.73)  
Correlations between difference and mean = -0.125

**Fig. 1.** Bland–Altman plot of agreement between the sIRLS and the IRLS applied in parallel during the study.

**Table 6**  
Prevalence of the WED/RLS symptoms, agreement, and correlation between the sIRLS and IRLS.

Items	Prevalence (%)			
	N	sIRLS	IRLS	$p^a$
1. Discomfort	164	95.1	93.0	0.22
2. Need to move	155	93.3	91.6	0.63
3. Relief	155	93.9	92.9	1.00
4. Sleep disturbance	155	84.8	80.0	0.18
5. During the day	155	79.9	75.5	0.38
6. RLS on the whole	155	93.3	93.6	1.00
7. How often	155	96.3	94.8	0.50
8. How severe	155	96.9	94.8	0.25
9. Daily activities	155	70.1	60.7	0.03 <sup>a</sup>
10. Mood disturbance	155	68.9	63.9	0.42
Total score	155	–	–	–

<sup>a</sup> McNemar's exact test.

(9.1). The results of the second sIRLS that was mailed in by the patients were similar in both English speaking countries and European countries: 15.7 (8.6) vs. 14.8 (8.5).

#### 4. Discussion

This study demonstrates that the sIRLS is an accurate and reliable self-completing scale to assess the severity of RLS symptoms. The sIRLS has been designed as a self-applied assessment with the same structure and components as the IRLS and with wording adjusted for its use by patients as opposed to clinicians. In the present study, the clinimetric attributes of the sIRLS were analyzed in parallel with the original IRLS. Data were captured in an independent manner to guarantee their reliability and validity.

As the sIRLS is derived from an already validated scale, the IRLS, the latter could be considered “the criterion” or “gold standard” against which the new instrument has to be tested. The results of the clinimetric attributes of the IRLS tested in the present study overlap those from previous validation studies of this scale [1,7,8]. This fact supports the consistency of the findings comparing the original IRLS and the new scale.

sIRLS data quality were excellent, with less than 0.5% missing data. Scores were properly distributed across the items, covering the theoretical score limits (Table 2), and floor and ceiling effects were irrelevant for the total score (Table 3). The parameters considered for scaling assumptions (mean, similar item variances, item-total correlations) were also appropriate for the sIRLS. These results, therefore, suggest that feasibility/acceptability and scaling assumptions of the sIRLS are adequate [16,17] and comparable to those of the IRLS.

The exploratory factor analysis showed that the sIRLS is a uni-dimensional scale with all items grouped around the construct (severity of the RLS), similar to the IRLS. All parameters related with the reliability of the sIRLS, both for internal consistency and stability, were clearly higher than the respective threshold values [15,18,21]. These are relevant outcomes related to the appropriateness of the scale, as well as to the meaning and reproducibility of its scores. Similar corresponding results were obtained with the IRLS.

In regard to the precision of the sIRLS, the SEM was lower than the value of ½ SD at baseline, suggesting adequate precision [22]. This statistic parameter depends on the sample variability and scale stability, keeps relatively constant independent of the sample, and reflects the standard deviation of the distribution of repeated questionnaire scores. Thus, the SEM represents a link between the reliability and the sensitivity to change of the measure [22–24].

The agreement between both scales may be considered an inter-rater reliability analysis (patient and doctor) and also a comparative analysis of the concurrence between measures (Tables 4 and 5). The results of this analysis showed that there is a “substantial” to “almost perfect” agreement between the respective components of both scales, with very few cases scoring out of the 95% CI for agreement established for the total scores [25]. Nonetheless, these results are referred to the collective of patients in the study considered as a whole. When the subjects and items are individually considered, a variable proportion of discrepancies in the scores is found, albeit most of them differing only by one point (Table 4). A similar situation occurs with the total score, as demonstrated by 95%CI of agreement (−5.50, 6.73) clearly higher than thresholds proposed for a minimal clinically significant difference (SEM, 3.48; ½ SD, 4.17) [26–28]. These findings have relevance for use of the scale in single patients.

The association between sIRLS and IRLS components was strong, with high correlation coefficients (Table 5). As expressed in the Data Analysis section, this outcome represents a modality of concurrent validity between an original instrument (IRLS) and the derived one (sIRLS). The high correspondence between the scales, reaching a 0.94 coefficient for the total score, is indicative of almost equivalent results with both scales.

This study verified the stability of the sIRLS as a measure of RLS symptom severity. This stability was maintained when patients from English speaking countries were compared to patients from European countries. The European group scored higher on both the sIRLS and the IRLS when completed in the office supporting the similarity between the two scales, even when symptom severity is different. The inclusion of several centers across the globe conveys both strength and limitation to this study. The strength is derived from the diversity of the patients included. The absence of validated translations for the sIRLS and the restriction to only English-speaking patients in the European group is a limitation of this study. This effect is mitigated by the use of English for all study related materials in all patients. Staff limitations, budget restrictions and limited availability of English speaking RLS patients delayed patient recruitment.

Longitudinal studies are needed to assess the sensitivity of the sIRLS to reflect medications changes or interventions. Another limitation of both scales is their design and validation as measures of symptom severity in already diagnosed RLS patients.

It is very convenient to have a valid instrument for self-evaluation of RLS severity. Neither the IRLS nor the sIRLS are diagnostic instruments for RLS. These severity scales should be used only in patients with an established diagnosis of RLS.

In conclusion, the sIRLS is a feasible, unidimensional, reliable, precise, and valid scale for self-evaluation of severity of RLS. The results from this first validation study demonstrate a close similarity with the IRLS from which the sIRLS is derived. These findings support the use of this self-administered form of the IRLS in the same broad range of clinical and non-clinical conditions for which the original rating scale was devised, validated and implemented.

## 5. Recommendations for the use of the scales

Because of some slight variability between the responses obtained between the sIRLS and the IRLS, we recommend that for serial measures of RLS severity over time that either the sIRLS alone or the IRLS alone be used as the only scale for all measurements. An exception would be when clarity over the questions is established initially with the IRLS as the first administration of the scale and when the sIRLS is employed in all subsequent administrations of the scale. Under these circumstances we recommend that statistical comparisons only be made only amongst the sIRLS versions of the scale.

## Acknowledgements

Limited financial support for this study was provided by the International Restless Legs Syndrome Study Group (IRLSSG). The current IRLSSG executive committee consists of Diego Garcia-Borreguero, Federica Provini, Rochelle Zak, Birgit Högl, Cornelius Bachman, Raffaele Ferri, Roselyne Rijsman, Denise Sharon, Garima Shukla, John Swieca, Arthur S. Walters, John W. Winkelman and Richard P. Allen.

The authors wish to thank Cathy Ann Goldstien, MD and Claudia Trenkwalder, MD for their RLS expertise and suggestions regarding this project. Special thanks to Karla Dzienkowski, Executive Director of the RLS Foundation, the RLS Foundation and their members for their participation, comments and input in the preliminary surveys detailing the instructions for completion of the sIRLS.

## Conflict of interest

The ICMJE Uniform Disclosure Form for Potential Conflicts of Interest associated with this article can be viewed by clicking on the following link: <https://doi.org/10.1016/j.sleep.2018.10.014>.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.sleep.2018.10.014>.

## References

- [1] Walters AS, LeBrocq C, Dhar A, et al., International Restless Legs Syndrome Study Group. Validation of the international Restless Legs Syndrome study group rating scale for Restless Legs Syndrome. *Sleep Med* 2003;4:121–32.
- [2] Bogan RK, Bornemann MAC, Kushida CA, et al., the XP060 Study Group. Long-term maintenance treatment of Restless Legs Syndrome with Gabapentin enacarbil: a randomized controlled study. *May Clin Proc* 2010;85(6):512–21. <https://doi.org/10.4065/mcp.2009.0700>.
- [3] Winkelman JW, Sethi KD, Kushida CA, et al. Efficacy and safety of pramipexole in Restless Legs Syndrome. *Neurology* 2006;67(6):1034–9. Epub 2006 Aug 23.
- [4] Hening WA, Allen RP, Ondo WG, et al., On behalf of the SP792 Study Group. Rotigotine improves Restless Legs Syndrome: a 6-month randomized, double-blind, placebo-controlled trial in the United States. *Mov Disord* 2010;25(11):1675–83.
- [5] Walters AS, Ondo WG, Dreykluft T, et al., On behalf of the TREAT RLS 2 (Therapy with Ropinirole: Efficacy and Tolerability in RLS 2) Study Group. 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(12):1414–23.
- [6] Trenkwalder C, Benes H, Grote L, et al., for the RELOXYN Study Group. Prolonged release oxycodone–naloxone for treatment of severe Restless Legs Syndrome after failure of previous treatment: a double-blind, randomized, placebo-controlled trial with an open-label extension. *Lancet Neurol* 2013;12(12):1141–50. [https://doi.org/10.1016/S1474-4422\(13\)70239-4](https://doi.org/10.1016/S1474-4422(13)70239-4). Epub 2013 Oct 18. Erratum in: *Lancet Neurol* 2013;12(12):1133.
- [7] Wunderlich GR, Evans KR, Sills T, et al., International Restless Legs Syndrome Study Group. An item response analysis of the international Restless Legs Syndrome study group rating scale for Restless Legs Syndrome. *Sleep Med* 2005;6:131–9.
- [8] Abetz L, Arbuckle R, Allen RP, et al. The reliability, validity and responsiveness of the International Restless Legs Syndrome Study Group rating scale and subscales in a clinical-trial setting. *Sleep Med* 2006;7:340–9.
- [9] 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(2):133–5.
- [10] Allen RP, Oertel W, Walters AS, et al. Relation of the international Restless Legs Syndrome study group rating scale with the clinical global impression severity scale, RLS-6 and Restless Legs Syndrome quality of life questionnaire. *Sleep Med* 2013;14:1375–80.
- [11] Walters AS, Frauscher B, Allen RP, et al. Stebbins GT and members of the MDS committee on rating scales. Review of severity rating scales for the Restless Legs Syndrome: critique and recommendations. *Mov Disord Clin Practice* 2014;1:317–24.
- [12] Kohnen R, Martinez-Martin P, Benes H, et al. Rating of daytime and night-time symptoms in RLS: validation of the RLS-6 scale of Restless Legs Syndrome/Willis Ekblom disease. *Sleep Med* 2016;20:116–22.
- [13] Allen RP, Picchietti DL, Garcia-Borreguero D, et al., International Restless Legs Syndrome Study Group. Restless Legs Syndrome/Willis-Ekblom disease

- diagnostic criteria: updated International Restless Legs Syndrome Study Group (IRLSSG) consensus criteria – history, rationale, description, and significance. *Sleep Med* 2014;15(8):860–73.
- [14] Hurst H, Bolton J. Assessing the clinical significance of change scores recorded on subjective outcome measures. *J Manipulative Physiol Ther* 2004;27(1): 26–35.
- [15] Smith SC, Lamping DL, Banerjee S, et al. Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technol Assess* 2005;9(10):1–93.
- [16] McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;4: 293–307.
- [17] van der Linden FA, Kragt JJ, Klein M, et al. Psychometric evaluation of the multiple sclerosis impact scale (MSIS-29) for proxy use. *J Neurol Neurosurg Psychiatry* 2005;76:1677–81.
- [18] Terwee C, Bot SDM, de Boer RM, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- [19] Eisen M, Ware Jr JE, Donald CA, et al. Measuring components of children's health status. *Med Care* 1979;17:902–21.
- [20] Beaton DE, Bombardier C, Katz JN, et al. A taxonomy for responsiveness. *J Clin Epidemiol* 2001;54:1204–17.
- [21] Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002;11:193–205.
- [22] Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life. The remarkable universality of half a standard deviation. *Med Care* 2003;41:582–92.
- [23] Wyrwich KW, Nienaber MA, Tierney WM, et al. Linking clinical relevance and statistical significance in evaluating intraindividual changes in health-related quality of life. *Med Care* 1999;37:469–78.
- [24] Wyrwich KW, Bullinger M, Aaronson N, et al., the Clinical Significance Consensus Meeting Group. Estimating clinically significant differences in quality of life outcomes. *Qual Life Res* 2005;14:285–95.
- [25] Sloan JA. Assessing the minimally clinically significant difference: scientific considerations, challenges and solutions. *COPD* 2005;2:57–62.
- [26] de Vet HC, Terwee CB, Ostelo RW, et al. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes* 2006;4:54.
- [27] Wyrwich KW, Norquist JM, Lenderking WR, et al. Industry advisory committee of international society for quality of life Research (ISOQOL). Methods for interpreting change over time in patient-reported outcome measures. *Qual Life Res* 2013;22:475–83.
- [28] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.