Sleep Medicine 16 (2015) 379–384

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Original Article

Measurement properties of the Minimal Insomnia Symptom Scale as an insomnia screening tool for adults and the elderly

Albert Westergren [a,*], Jan-Erik Broman [b], Amanda Hellström [c,d], Cecilia Fagerström [c], Ania Willman [c,e], Peter Hagell [a]

[a] The PRO-CARE Group, School of Health and Society, Kristianstad University, Kristianstad, Sweden
[b] Department of Neuroscience, Psychiatry, Uppsala University, Uppsala, Sweden
[c] Department of Health, Blekinge Institute of Technology, Karlskrona, Sweden
[d] Department of Health and Caring Sciences, Linnæus University, Kalmar, Sweden
[e] The Department of Care Science, Faculty of Health and Society, Malmö University, Malmö, Sweden

ARTICLE INFO

Article history:
Received 10 June 2014
Received in revised form 3 October 2014
Accepted 9 October 2014
Available online 15 January 2015

Keywords:
MISS
Psychometric properties
Clinical insomnia
Invariant comparisons
Targeting
Reliability

ABSTRACT

Background: The psychometric properties of the three-item Minimal Insomnia Symptom Scale (MISS) were evaluated using the classical test theory. Different cut-offs for identifying insomnia were suggested in two age groups (≥6 and ≥7 among adult and elderly people, respectively). The aim of the present study was to test the measurement properties of the MISS using the Rasch measurement model, with special emphasis on differential item functioning by gender and age.
Methods: Cross-sectional MISS data from adult (age 20–64 years, $n = 1075$) and elderly (age 65+, $n = 548$) populations were analysed using the Rasch measurement model.
Results: Data generally met Rasch model requirements and the scale could separate between two distinct groups of people. Differential item functioning was found by age but not gender. The difference between the adult and elderly samples was lower for the originally recommended ≥6 points cut-off (0.09 logits) than for the ≥7 points cut-off (0.23 logits), but greater at the lower and higher ends of the scale.
Conclusions: This study provides general support for the measurement properties of the MISS. Caution should be exercised in comparing raw MISS scores between age groups, but applying a ≥6 cut-off appears to allow for valid comparisons between adults and the elderly regarding the presence of insomnia. Nevertheless, additional studies are needed to determine the clinically optimal cut-score for identification of insomnia.

1. Introduction

The Minimal Insomnia Symptom Scale (MISS) has been proposed as a brief insomnia-screening questionnaire. Its basic psychometric properties have been evaluated in two studies using the classical test theory (CTT) among general adult (age 20–64) and elderly (age 65+) populations [1,2]. In these studies, the MISS was able to identify adults, as well as elderly people, with and without clinical insomnia, according to International Classification of Diseases 10th revision (ICD-10) criteria and three items from the 25-item Uppsala Sleep Inventory as proxy gold standards, respectively [1,2]. However, the optimal cut-off was found to be lower among adults than in the elderly, which might suggest differences in the psychometric properties of MISS in the two age groups.

The difference in cut-offs was hypothesised to be related to variations in the sample size and age of the respective samples, as well as to the use of different gold standards [2]. It was further hypothesised that elderly people may perceive sleeping difficulties and related daytime impairments differently from adult people. That is, age might influence the experience of daytime impairments relating to poor sleep [2]. It is unknown whether gender influences the perception of sleeping difficulties and related daytime impairments, but insomnia has been found to be more common among women [3]. Thus, for the same insomnia level, the expected item response from different gender or age groups might vary [4]. This phenomenon is known as differential item functioning (DIF) and is problematic because valid quantitative comparisons assume that items work in the same way across different groups of individuals, such as age groups and genders [4]. The absence of DIF means that item responses depend on levels of insomnia, and not on the respondent's gender or age; otherwise (in the presence of DIF), scores may be biased [5,6].

The Rasch model allows for rigorous testing of measurement instruments and makes it possible to examine aspects of measurement beyond those offered by CTT, including detailed DIF analyses [7]. Rasch analysis is therefore considered to be superior to CTT for determining rating scale measurement properties [8].

* Corresponding author. School of Health and Society, Kristianstad University, SE-291 88 Kristianstad, Sweden. Tel.: +46 44 208550; fax: +46 44 204019.
E-mail address: Albert.Westergren@hkr.se (A. Westergren).

http://dx.doi.org/10.1016/j.sleep.2014.10.016
1389-9457/

The aim of the present study was to examine the measurement properties of the MISS using the Rasch measurement model, with a special focus on DIF by age and gender.

2. Methods

2.1. Samples

The present study was based on two samples: adult (20–64 years) and elderly (65+ years) [1,2]. The adult sample was randomly selected from the general Swedish population ($n = 1075$, 78% response rate, 555 females) [1]. The elderly sample was taken from one participating region in a national Swedish health-related survey of the elderly ($n = 548$, 61% response rate, 315 females) [2]. Both studies had ethical approval.

2.2. Instrument

The MISS [1] is an insomnia-screening tool consisting of three items, with five response categories (no, minor, moderate, severe and very severe problems), which are scored 0–4, respectively. This yields a total score ranging from 0 to 12 (12 = more insomnia). Previous studies have suggested cut-offs of ≥6 among adults [1] and ≥7 among the elderly [2].

2.3. The Rasch model

The Rasch model mathematically defines the requirements from item responses in order to yield linear measures [9]. According to the Rasch model, the probability of a certain item response is a function of the difference between the level of the measured construct (eg, insomnia) represented by the item and that possessed by the person. The model separately locates people and items on a common interval level logit (log-odd units) metric, with mean item location set at 0. If data accord with the model, linear measurement and invariant comparisons are possible [6,8,10,11].

2.4. Analysis

Rasch analysis was performed using RUMM2030 [12]. Two-tailed *p*-values <0.05, following Bonferroni adjustment, were considered to be significant. Since the calculation of a total MISS score and application of cut-offs for identification of cases with insomnia require complete item responses, analyses were based on respondents without any missing item response data ($n = 1542$).

2.5. Targeting, precision and reliability

Good targeting means that the scale represents the levels of insomnia reported by the sample, which is a prerequisite for reasonable assessments of the scale. Furthermore, poor targeting compromises measurement and precision, and therefore the ability to differentiate between people [6]. One indicator of targeting is the mean person location relative to the mean item location (ie, 0 logits) [6]. In general, mean person locations within ±0.5 logits suggest acceptable targeting [13,14]. However, with screening instruments, the greatest concern is to maximise precision and information around the cut-off point for case identification. This can be assessed by the information function (IF), which is the inverse of the measurement error and should therefore be greatest around the cut-off point. Precision also relates to the ability to separate a sample into distinct groups. This is reflected through score reliability. Reliability was estimated by the Person Separation Index (PSI), which is analogous to coefficient alpha [15] and can be used to derive the number of strata (ie, statistically distinct groups separated by ≥3 standard errors) that can be distinguished by the scale [16–18]. At least two strata would be required for a screening tool such as the MISS.

2.6. Response category functioning

Studying response category thresholds (ie, the locations where there is a 50/50 probability of responding in either of two adjacent categories) assesses whether response categories function as expected. Ordered thresholds mean that response categories appear in an expected order from less to more, whereas disordered thresholds mean that the location at which there is a 50/50 probability of responding (eg, either ‘moderate’ or ‘severe’) is at a lower level than that where there is a 50/50 probability of responding in either ‘minor’ or ‘moderate’ [6].

2.7. Model fit

Fit refers to the extent to which observed item responses accord with the Rasch model; that is, the extent to which they are predicted by, or recovered from, the model [8]. Model fit was examined graphically and statistically. Plots of observed item responses against model expectations (item characteristic curves, ICCs) were inspected. Empirical observations that accord with expectations support model fit. Standardised item fit residuals represent the discrepancies between observed and expected responses [8]. In general, individual item fit residuals should range between –2.5 and 2.5, with the ideal being 0 [19]. Negative values suggest local dependency (which is also implied by correlations >0.3 among item fit residuals), while positive values imply multidimensionality [8]. Sufficient fit to the Rasch model, thus, suggests that the basic assumptions of local independence and unidimensionality are supported. For a well-fitting scale, the mean standardised item fit residuals should be close to 0, with a standard deviation close to 1. Formally, the difference between expected and observed item response data is tested by the Chi-squared statistic, which should be non-significant to support model fit. The fit of respondents is also assessed according to the same principle as for items. Person misfit indicates unlikely responses, which may compromise item model fit. In the present study, mis-fitting people were therefore deleted in order to assess their impact on item fit.

While model fit is fundamental, its statistical assessment is complicated by the fact that the type I error rate increases with increasing sample size. It has therefore been recommended to adjust sample size when large data sets are analysed [8,19]. Sample size was therefore algebraically adjusted to 500 in the calculation of *p*-values [8], while leaving all other aspects of data (eg, locations, fit residuals) unaltered [19].

2.8. Differential item functioning

Differential item functioning (DIF) concerns whether items have the same meaning in different subgroups of respondents (eg, age and gender groups) and can result in measurement bias that challenges valid subgroup comparisons [6]. In the present study, DIF was tested by a two-way ANOVA of the residuals across insomnia levels (subgroups of people with similar MISS scores) for age (<65 vs 65+) and gender. In the case of uniform DIF (ie, a systematic difference between subgroups across insomnia levels), this was adjusted for by splitting the affected item into two new subgroup-specific items [6,20]. For example, an item exhibiting DIF by age would be split and analysed as two separate items, one for each age group (and analogously for men and women in the case of DIF by gender). This procedure is valid because the Rasch model can accommodate missing data [4,19,20].

The clinical significance of DIF was studied by testing whether the same total scores reflected the same levels of insomnia across