Journal of Psychosomatic Research 87 (2016) 85–92

Contents lists available at ScienceDirect

Journal of Psychosomatic Research

A meta-analysis of diagnostic accuracy of three screening tools for insomnia

Hsiao-Yean Chiu [a], Ling-Yin Chang [b], Yu-Jung Hsieh [a], Pei-Shan Tsai [a,c,*]

[a] School of Nursing, College of Nursing, Taipei Medical University, Taipei, Taiwan
[b] Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan.
[c] Sleep Science Center, Taipei Medical University Hospital, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 29 March 2016
Received in revised form 23 June 2016
Accepted 24 June 2016

Keywords:
Diagnostic metaanalysis
Insomnia
Insomnia screening
Instrument validation

ABSTRACT

Background: Insomnia is a highly prevalent health complaint in the modern societies; however, insomnia remains under-diagnosed and under-treated. Although screening tools, including the Insomnia Severity Index (ISI), Athens Insomnia Scale (AIS), and Pittsburg Sleep Quality Index (PSQI), are widely used for assessing the risk of insomnia, the diagnostic properties have yet to be summarized in a systematic manner.
Objectives: To estimate and to compare the diagnostic accuracy of the ISI, AIS, and PSQI for insomnia screening.
Data sources: We systematically searched EMBASE, PubMed, PsycINFO, CINAHL and Chinese Electronic Periodic Services for data from their inception to May 20, 2015.
Data selection: Original articles that had assessed the sensitivity and specificity of the ISI, AIS, or PSQI against a reference standard in adult participants (age > 18) were included.
Results: A total of 19 studies comprising 4693 participants were included. The pooled sensitivity for the ISI, AIS, and PSQI was 88% (95% confidence interval [CI] = 0.79 to 0.93), 91% (0.87 to 0.93), and 94% (0.86 to 0.98), respectively. The pooled specificity was 85% (0.68 to 0.94), 87% (0.68 to 0.95), and 76% (0.64 to 0.85); and the pooled DORs was 41.93 (8.77 to 200.33), 67.7 (23.4 to 196.1), and 53 (15.5 to 186.2), respectively. The summary estimates did not differ significantly among the ISI, AIS and PSQI (all $P > 0.05$).
Conclusions: The current evidence indicates that the ISI, AIS, and PSQI yield comparable diagnostic properties for insomnia screening.

1. Introduction

Insomnia is a highly prevalent health complaint, with a prevalence rate ranging from 6% to 34.5% [1–4] depending on the differences in populations and diagnostic criteria used. Insomnia is associated with increased healthcare utilisation [5], work productivity loss [6], cognitive functions impairment [7], and reduced quality of life [8]. Insomnia remains an underdiagnosed and undertreated health problem [9]despite its high prevalence and the substantial negative consequences [9–12].

Polysomnography is the gold standard for identifying sleep disorders, such as sleep apnoea and periodic limb movement disorder; however, it is not recommended as a routine evaluation for insomnia [13]. Although structured or semistructured clinical interviews are widely adopted for establishing a clinical diagnosis of insomnia [14], it is impractical for routine clinical use because it is time consuming and requires well-trained practitioners. Therefore, for wider practicability, instruments that are brief, valid, reliable and easy-to-use are essential for clinical and community insomnia assessment.

A diagnosis of insomnia largely relies on standard diagnostic criteria for insomnia (e.g., the International Statistical Classification of Diseases and Related Health Problems [ICD], the Diagnostic and Statistical Manual of Mental Disorders [DSM], and International Classification of Sleep Disorders [ICSD]). A useful and accurate screening tool for insomnia should therefore be established according to these diagnostic criteria. Two instruments, the Insomnia Severity Index (ISI) and the Athens Insomnia Scale (AIS), have been developed according to standard insomnia diagnostic criteria. The ISI captures the diagnostic criteria for insomnia outlined in the DSM-IV and ICSD [15], and the AIS was designed for quantifying sleep difficulty based on the ICD-10 [16]. Of note, an expert panel of sleep researchers recommended the Pittsburgh Sleep Quality Index (PSQI) as a standard assessment tool of insomnia [17] although it was not originally designed for use in assessing the risk of insomnia. These three instruments have been widely used in research fields (e.g., treatment efficacy assessment) and clinical settings [18]. Understanding the diagnostic accuracy and properties among these three instruments might assist clinicians and researchers in

* Corresponding author at: 250 Wu-Hsing St., Taipei 110, Taiwan.
E-mail address: ptsai@tmu.edu.tw (P.-S. Tsai).

http://dx.doi.org/10.1016/j.jpsychores.2016.06.010
0022-3999/

selecting appropriate instruments for use in screening patients for suspected insomnia.

Thus far, only one study assessed and compared the diagnostic accuracy of the ISI and PSQI for insomnia in adults. That study consisting of 79 patients with low back pain [19] examined discriminatory properties of the PSQI and ISI and found that the PSQI and ISI achieved comparable accuracy in insomnia screening. However, the study included a small sample size and thus lacked sufficient statistical power to detect differences. In addition, it merely focused on the specific population which limited its generalizability. While previous research has neither compared the diagnostic properties among these three instruments nor has examined whether the PSQI is also an appropriate screening tool for insomnia compared with other instruments that were developed based on standard insomnia diagnostic criteria (i.e., the ISI and AIS), conducting a diagnostic metaanalysis to evaluate the diagnostic properties for insomnia screening among these three instruments is a major research priority.

This diagnostic meta-analysis was performed to estimate and compare the diagnostic accuracy of the ISI, AIS and PSQI for insomnia screening. Results from the present study could facilitate healthcare providers to select an appropriate instrument in insomnia screening in clinical settings and research fields.

2. Materials and methods

This meta-analysis was conducted according to PRISMA statement, which provided a detailed guideline of preferred reporting items for systematic review and meta-analysis [20].

2.1. Identification of studies

A comprehensive search for original studies on the diagnostic accuracy of the ISI, AIS and PSQI for insomnia screening was performed on EMBASE, PubMed, PsycINFO, CINAHL and Chinese Electronic Periodic Services for data from their inception to May 20, 2015. The search terms used were as follows: (Insomnia Severity Index OR Athens Insomnia Scale OR Pittsburgh Sleep Quality Index) AND (sensitivity OR specificity OR validity OR reliability OR validation OR cutoff value). The references from the identified studies and relevant published reports were manually searched to identify studies eligible for inclusion in our review.

2.2. Inclusion criteria and study selection

We included original articles that had assessed the sensitivity and specificity of the ISI, AIS, or PSQI in comparison with reference standard test results in adult participants (age > 18). The studies should be available as a full publication in a peer-reviewed journal. Thesis and dissertation with full-text was also included for further analyses to avoid the possibility of publication bias. The titles and abstracts of eligible articles identified through the aforementioned search criteria were independently screened by two reviewers (HYC and YJH). Full-text articles were reviewed after discarding duplicates from the potentially eligible articles to determine whether they met the inclusion criteria.

2.3. Data extraction

Data, including authors, year of publication, country, study design, populations and controls, age, percentage of females, number of participants, reference test, measurement tool, cutoff value, were independently extracted from each publication by the two reviewers using predesigned data extraction form. When more than one pair of sensitivity and specificity were reported according to various cutoff values, we chose the cutoff value that produced the highest Youden's index [21]. Disagreements were resolved through discussion.

2.4. Assessments of methodological quality of studies

The two reviewers (HYC and YJH) individually assessed the quality of each included study according to the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2), a systematic review checklist recommended by the Cochrane Collaboration [22]. QUADAS-2 evaluates the risk of bias and concerns of applicability for three domains (patient selection, index test, reference standard) and the risk of bias in one domain (study flow and timing). Each domain contains a set of signalling questions to help the reviewers reach the judgments regarding bias and applicability. Reviewers' responses were rated as “yes”, “no”, and “unclear”. Domains with at least one “no” response and those with “yes” response in the entire domain were rated high and low risk of bias, respectively. The “unclear” risk of bias was rated when insufficient data were reported to permit a judgment.

2.5. Data analysis

Data were analysed using Stata, version 14.0 (Stata Corp LP, College Station, Texas, USA) with Midas and Metandi user-written commands, SAS version 9.0.2 (SAS Institute Inc., Cary, North Carolina, USA)with Proc Mixed module, and Review Manager 5.3. We performed a bivariate diagnostic meta-analysis using a generalised linear mixed model [23] to estimate pooled sensitivity, specificity, positive and negative likelihood ratios and summary diagnostic odds ratio (DOR). DOR refers to the ratio of the odds of positivity in disease relative to the odds of positivity in the non-disease, which combines the strengths of sensitivity and specificity [24]. If the included study had zero cells, 0.5 was added to each cell in the underlying 2×2 tables to avoid problems associated with sensitivity or specificity equalling 1 [25]. I^2 Statistic describing the percentage of total variation across studies resulting from heterogeneity rather than chance was used to assess between-study heterogeneity in terms of sensitivity and specificity [26]. An approximate I^2value of 0% indicates no observed heterogeneity and values higher than 50% represent a substantial heterogeneity. The priori causes of significant between-study heterogeneity among the included studies were explored by investigating the threshold effect through Spearman correlation analyses and adding covariates into moderator analyses. The covariates were demographic data, such as age and percentage of women, clinical characteristics, such as comorbidity, and methodological features, such as the study design and location (country) of the study. Moderator analyses were limited to groups represented by at least two studies to ensure sufficient data for analysis. Publication bias was evaluated using Deek's funnel plots. Funnel plot asymmetry was detected using a regression test of diagnostic log odds ratio against 1/sqrt (effective sample size), weighed by effective sample size [27]; $P < 0.10$ indicated significant asymmetry for the slope coefficient.

3. Results

3.1. Search results

The review (workflow illustrated in Fig. 1) includes 1252 abstracts retrieved from the electronic database search, of which 711 duplicates were discarded using the Endnote software. Following initial screening, 519 articles with unrelated content, studies published in Turkish or Polish languages, unexamined sensitivity and specificity and studies with children or adolescents as participants were excluded. Among the remaining articles, 22 were considered potentially suitable. Furthermore, following a review of