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

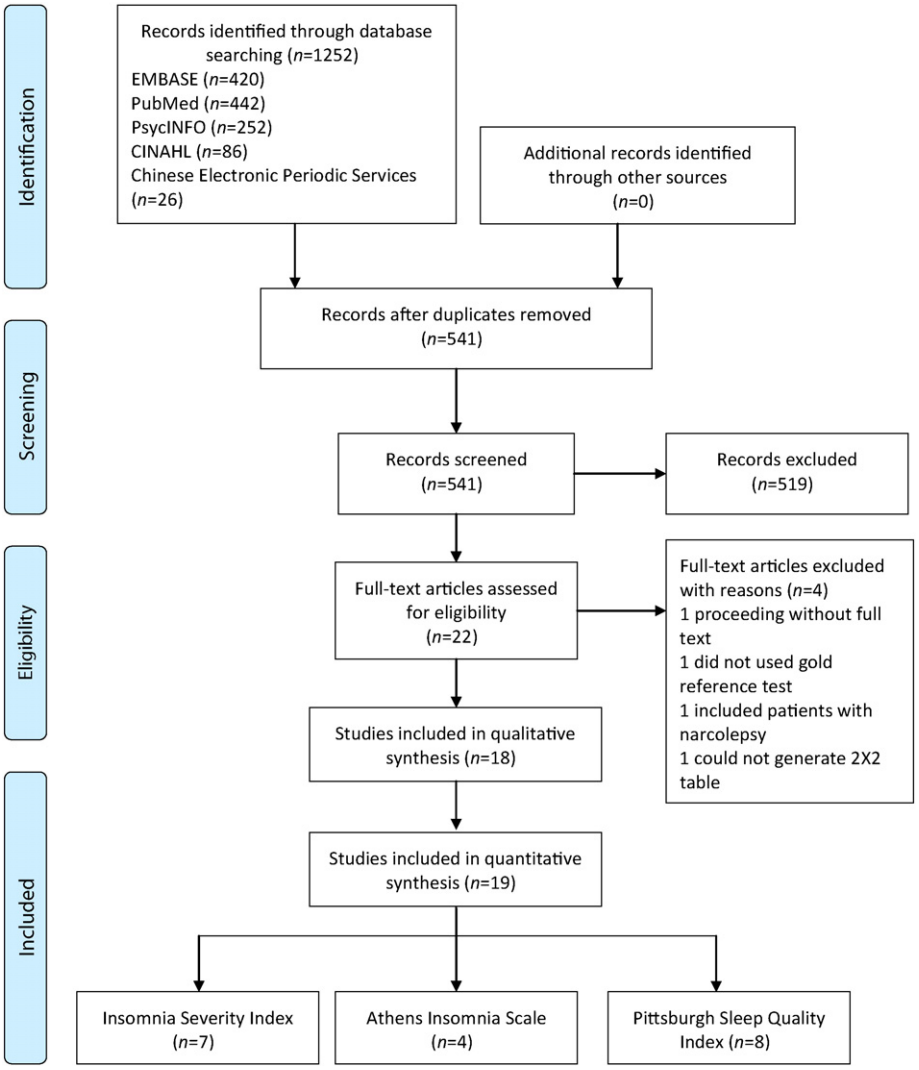

Fig. 1. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2009 flow diagram.

the full text, four studies were excluded: the full text for one study was not provided by the original author study [28] had not used gold standard tests, such as DSM, ICD and ICSD as the reference tests [29], one study directly considered patients with insomnia and those with narcolepsy as poor sleepers without accurately diagnosing insomnia [30] and one study could not generate a 2 × 2 table [31]. Altogether, 18 studies met the inclusion criteria were selected for further analyses [19,32–48]; 19 studies were used for quantitative analyses because one study evaluated both the ISI and PSQI [19].

3.2. Study characteristics

Of the 19 studies, seven, four and eight studies used the ISI, AIS, and PSQI as the index test, respectively. Details regarding demographic data and clinical characteristics of these studies are presented in Table 1. Among these studies, eight studies [19,36,38,40,41,46,48] evaluating the ISI included 2473 participants with a mean age of 47.2 years. Three studies used a case-control design and one employed a retrospective cohort design. The International Classification of Sleep Disorders (ICSD-II) was the most commonly used reference standard for insomnia diagnosis (k = 3), with cutoff values ranging from 9 to 15.5 among the included studies. Four studies [37,39,42,43] evaluating the AIS included 779 participants with a mean age of 47.2 years. Of these studies, three used a cross-sectional design. The cutoff values of the AIS for insomnia were 6 and 6.5. Eight studies [19,32–35,44,45,47] evaluating the PSQI included 1441 participants with a mean age of 41.1 years. Of these studies, four used a case-control design. The DSM-IV was the most frequently used reference test for insomnia diagnosis (k = 6), with cutoff values ranging from 5 to 8 among these studies.

3.3. Quality of studies

The QUADAS-2 quality assessment of the included studies is presented in Table 2. The quality of the methodology employed in these studies was unsatisfactory. In four of the seven studies investigating the accuracy of the ISI, the risk of bias for patient selection was high because of the use of a case-control study design. The index test results and the reference standard results were neither interpreted in a blinded fashion nor in an appropriate order in most of studies. Half of studies were rated as unclear risk of bias in the domain of flow and timing.

In the studies investigating the accuracy of the AIS, two studies were rated as high risk of bias for patient selection because they employed a case-control study design. In most studies that did not report blinding of the assessors when interpreting the index test or reference standard,

Table 1
Characteristics of included studies in investigating Insomnia Severity Index, Athens Insomnia Scale, and Pittsburgh Sleep Quality Index.

First author, year	Country	Study design	Population	Control	Mean age	Female (%)	Sample size	Measure	Reference test	Clinical interviewer for classifying insomnia	Cut-off value
Alsaadi, 2013 [16]	Australia	Cross-sectional	Patients with low back pain	None	43.9	51	79	ISI	ICSD-II	No mention	14
Cho, 2014 [45]	Korea	Case-control	Chronic primary insomniacs and comorbid insomniacs	OSA	53.9	62.6	302	ISI	DSM-IV-TR	No mention	15.5
Gagnon, 2013 [33]	Canada	Cross-sectional	Outpatients with psychiatric and medical disorders	None	49.0	65.3	101	ISI	DSM-IV + ICSD2 + DSM-V	Semistructure interview by 1 trained research assistant with a BA in psychology and 5 clinical psychology graduate students.	14
Morin, 2011 [35]	Canada	Case-control	Primary insomniacs	Healthy subjects	50.7	61.2	245	ISI	DSM-IV	Semistructure interview but mention who performed the interviews.	11
Savard, 2005 [37]	Canada	Cross-sectional	Patients with breast cancer	None	UK	100	210	ISI	DSM-IV + ICSD-1	Insomnia Interview Schedule was done by graduate students in clinical psychology, who were supervised by a certificated psychologist.	12
Severson, 2013 [38]	Canada	Retrospective cohort	Primary insomniacs	SDs and psychiatric disorders	45.4	56.8	1207	ISI	ICSD-II	Two board-certificated sleep physicians reviewed all patients' charts.	15
Yang, 2009 [43]	Taiwan	Case-control	Primary insomniacs	Healthy subjects	40.4	57.1	329	ISI	ICSD-II	Clinical interviews were administrated by trained clinical psychologists or graduate students.	9
Jeong, 2015 [34]	Korea	Cross-sectional	Firefighter and rescue workers	None	40.3	6.3	221	AIS	ICD-10, DSM-IV-TR, Research Diagnostic Criteria	Structure Clinical Interview was administrated by clinicians trained by a board-certified psychiatrist.	6
Okajima, 2013 [36]	Japan	Case-control	Chronic insomnia outpatients (primary + comorbid insomnia)	Subjects with PSQI score < 6	48.8	58	640	AIS	ICSD-II	No mention.	6.5
Soldatos, 2003 [39]	Greece	Cross-sectional	Primary insomniacs, psychiatric outpatients and inpatients, healthy patients	None	42.9	58	299	AIS	ICD-10	Nonstructural Clinical Interviews was administrated by two co-investigators	6
Sun, 2011 [40]	Taiwan	Cross-sectional	Cancer patients	None	56.7	59.5	195	AIS	DSM-IV	No mention.	6.5
Aloba, 2007 [29]	Nigeria	Cross-sectional	Undergraduate and graduate students	None	23.2	38.8	520	PSQI	DSM-IV + ICSD-R	Insomnia diagnosis was administrated by four trained psychiatrists	5
Alsaadi, 2013 [16]	Australia	Cross-sectional	Patients with low back pain	None	43.9	51	79	PSQI	ICSD-II	No mention.	6
Backhaus, 2002 [30]	Germany	Case-control	Primary insomniacs	Healthy subjects	44.8	54.9	125	PSQI	DSM-IV	No mention.	5
Doi, 2000 [31]	Japan	Case-control	Primary insomniacs	Healthy subjects	50.4	47.9	96	PSQI	DSM-IV	Clinical Interviews were administrated by trained psychiatrists.	5.5
Farrahi Moghaddam, 2012 [44]	Iran	Case-control	Patients with psychiatric diseases	Healthy subjects	35.5	55.0	258	PSQI	DSM-IV	Clinical Interviews were administrated by psychiatrists.	6
Fictenberg, 2001 [32]	USA	Cross-sectional	Patients with TBI	None	33.8	41	50	PSQI	DSM-IV	No mention.	8
Tsai, 2005 [41]	Taiwan	Case-control	Primary insomniacs	Healthy subjects	39	63.9	208	PSQI	DSM-IV	Clinical Interviews were administrated by a general practitioner and a psychiatrist.	6
Tzeng, 2012 [42]	Taiwan	Cross-sectional	Patients with cancer	None	58.3	34.1	205	PSQI	DSM-IV	Did not mention who performed structure Interviews.	8

ICSD-II = Second edition of the International Classification of Sleep Disorders. IDI = insomnia Diagnostic Interview (IDI); DIS = Diagnosis of Clinical Schedule. IIS = Insomnia Interview Schedule. SDs = sleep disorders. OSA = obstructive sleep apnea. PSQI = Pittsburgh Sleep Quality Index. ICD = International Statistical Classification of Diseases and Related Health Problems 10th Revision. DSM = the Diagnostic and Statistical Manual of Mental Disorders, 4th. Edition. ICSD-R = The International Classification of Sleep Disorders, Revised. ICSD-II = The International Classification of Sleep Disorders, 2nd edition. UK = unknown.

the risk of bias for the index test and reference standard remained unclear. Three studies were rated as unclear risk of bias in the domain of flow and timing.

Regarding the studies investigating the accuracy of the PSQI, four studies were rated as high risk of bias for patient selection because they employed a case-control study design. In four studies the determination

Table 2
QUADAS-2 risk of bias assessment.

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
ISI							
Alsaadi, 2013 [16]	L	L	U	U	L	L	L
Cho, 2014 [45]	H	U	L	U	L	L	L
Gagnon, 2013 [33]	L	L	L	H	L	L	U
Morin, 2011 [35]	H	U	U	L	L	L	L
Savard, 2005 [37]	H	L	U	U	L	L	L
Severson, 2013 [38]	L	U	L	H	L	L	L
Yang, 2009 [43]	H	U	L	U	L	L	L
AIS							
Jeong, 2015 [34]	L	U	U	U	L	L	L
Okajima, 2013 [36]	H	U	U	U	L	L	L
Soldatos, 2003 [39]	H	L	L	L	L	L	L
Sun, 2011 [40]	L	U	U	U	L	L	L
PSQI							
Aloba, 2007 [29]	L	L	U	U	L	L	L
Alsaadi, 2013 [16]	L	L	L	U	L	L	L
Backhaus, 2002 [30]	H	H	L	H	L	L	L
Doi, 2000 [31]	H	H	L	H	L	L	L
Farrahi Moghaddam, 2012 [44]	H	H	L	U	L	U	L
Fictenberg, 2001 [32]	L	L	U	U	L	L	L
Tsai, 2005 [41]	H	H	L	U	L	L	L
Tzeng, 2012 [42]	L	U	L	U	L	L	L

QUADAS-2, the revised Quality Assessment of Diagnostic Accuracy Studies. L, low. H, high. U, unclear. ISI, Insomnia Severity Index. AIS, Athens Insomnia Scale. PSQI, Pittsburgh Sleep Quality Index.

of the index test level was not conducted in a way in which assessors interpreted the results without the knowledge of the reference test results. Most studies (k = 6) interpreted the reference standard without the knowledge of the index test results. The majority of studies (k = 6) were rated as unclear risk of bias in the domain of flow and timing.

The applicability of the ISI, AIS and PSQI was not a major concern in any of the included studies.

3.4. The summary diagnostic accuracy of the ISI, AIS, and PSQI

The paired sensitivity and specificity with 95% confidence intervals of each study in the forest plot are presented in Fig. 2 and Table 1. Specifically for the ISI, we found a pooled sensitivity of 88% (95% CI = 0.79 to 0.93), a pooled specificity of 85% (95% CI = 0.68 to 0.94), a positive likelihood ratio of 6.01, a negative likelihood ratio of 0.15, and a pooled

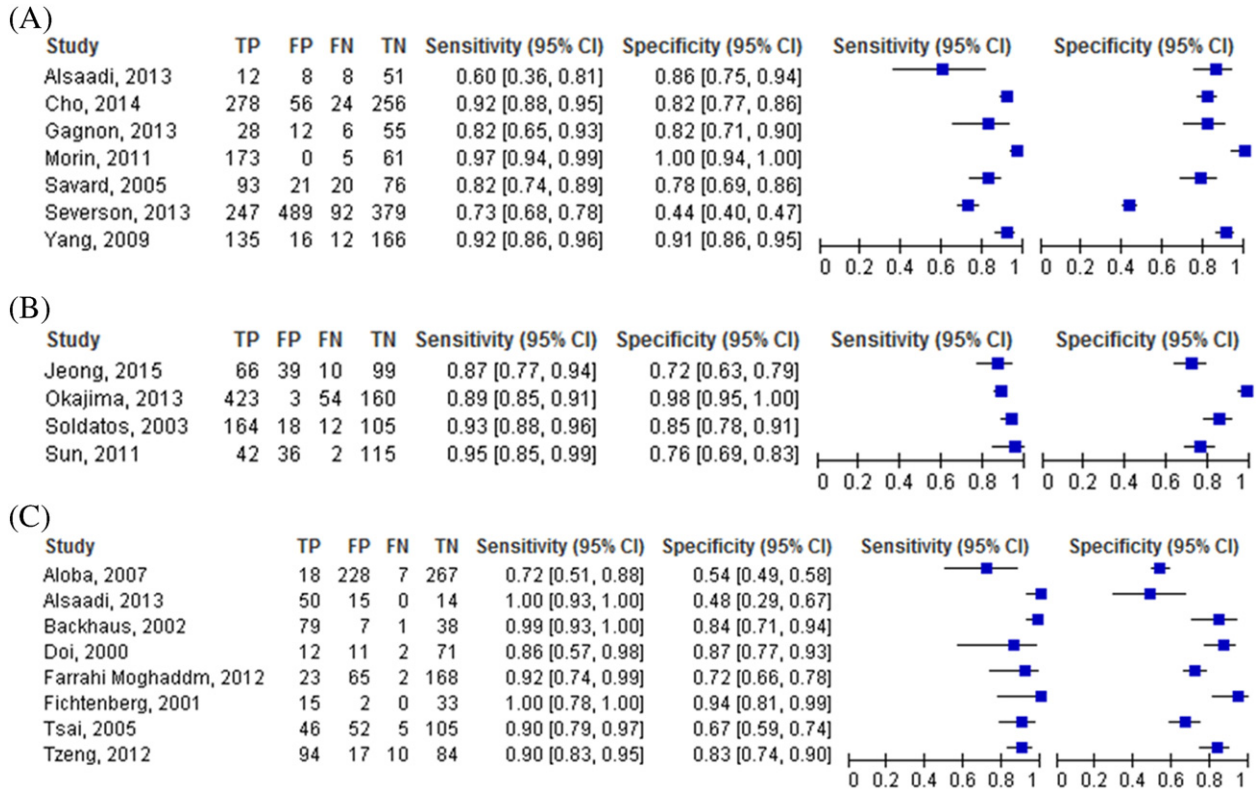

Fig. 2. Forest plots of pair of sensitivity and specificity in each study included for (A) Insomnia Severity Index, (B) Athens Insomnia Scale, and (C) Pittsburgh Sleep Quality Index.

DOR of 41.93 (95% CI = 8.77 to 200.33). In terms of the AIS, we found a pooled sensitivity of 91% (95% CI = 0.87 to 0.93), a pooled specificity of 87% (95% CI = 0.68 to 0.95), a positive likelihood ratio of 7.04, a negative likelihood ratio of 0.10, and a pooled DOR of 67.7 (95% CI = 23.4 to 196.1). With regard to the PSQI, we found a pooled sensitivity of 94% (95% CI = 0.86 to 0.98), a pooled specificity of 76% (95% CI = 0.64 to 0.85), a positive likelihood ratio of 3.94, a negative likelihood ratio of 0.07, and a pooled DOR of 53 (95% CI = 15.5 to 186.2).

3.5. Heterogeneity of diagnostic accuracy of the ISI, AIS, and PSQI

Between-study heterogeneity of pooled sensitivity were identified in the ISI, AIS and PSQI (ISI: Q = 182.1, $P < 0.001$; AIS: Q = 7.59, $P = 0.06$; PSQI: Q = 30.7, $P < 0.001$), with over half of the variance attributable to heterogeneity across the studies (I^2 = 97.0%, 60.5% and 77.2%, respectively) through chi-square and I^2 statistics. The heterogeneity for pooled specificity were confirmed for the ISI, AIS and PSQI (Q = 789.9, 46.7 and 189.9, respectively; all $P < 0.001$), with over half of the variance attributable to heterogeneity across studies (I^2 = 99.2%, 93.6% and 96.3%, respectively).

3.6. Threshold effect tests and moderator analyses for heterogeneity

We explored the causes of heterogeneity by investigating the threshold effects and performing predefined moderator analyses. In terms of the threshold effect, nonsignificant correlations between the sensitivity and specificity estimates were determined for the ISI, AIS and PSQI ($P = 0.3$, 0.8 and 0.9, respectively) through Spearman correlation analyses.

The moderator analyses for the pooled sensitivity and specificity of the ISI revealed that the studies using a case-control design yielded larger sensitivity and specificity compared with those using cross-sectional and retrospective cohort study designs (sensitivity: 94% vs. 79%, $P < 0.001$; specificity: 94% vs. 74%, $P = 0.04$; Table 4). Regarding the PSQI, studies that used the DSM-IV as the reference test had a higher specificity compared with those that used the ICSD-II and a combination of the DSM-IV and ICSD-R (81% vs. 54%, $P = 0.008$).

3.7. Comparisons of summary diagnostic accuracy among the ISI, AIS, and PSQI

A significant difference was not observed in the pooled sensitivity, specificity and DOR among the ISI, AIS and PSQI (all $P > 0.05$) (Table 3).

3.8. Publication bias

Deeks' funnel plot asymmetry tests in the ISI, AIS and PSQI did not reveal a significant difference ($P = 0.25$, 0.19 and 0.66, respectively) for the slopes coefficient, suggesting data symmetry and a relative low likelihood of publication bias.

4. Discussion

According to our review of relevant literature, this is the first diagnostic meta-analysis summarizing and comparing the diagnostic properties of the ISI, AIS and PSQI using DSM-IV, ICSD-II and ICD-10 as the reference standards, for identifying individuals with insomnia. The results of the current meta-analysis revealed that the ISI, AIS and PSQI are comparable for diagnostic accuracy as evidenced by similar sensitivity, specificity, and DOR for insomnia. The findings are reliable and applicable to common settings where patients with insomnia are managed because the current meta-analysis included individuals from both clinical and community settings and the data were retrieved from studies conducted in various countries across several continents.

Our findings revealed that both the ISI and PSQI are able to discriminate individuals with insomnia from those without insomnia although the presence of comorbid conditions can interfere with diagnostic accuracy of a test [49]. Therefore, both the ISI and PSQI can be used for screening individuals with insomnia regardless of the underlying cause of insomnia (primary and secondary insomnia).

Consistent with a comparative study [50] using the DSM-IV, ICD-10, and ICSD-II to evaluate insomnia symptoms in the American Insomnia Survey, our findings showed that studies employing the ICSD-II as the reference test had lower pooled sensitivity in the ISI and specificity in the PSQI compared with other reference tests. The inclusion of different diagnostic criteria of insomnia is likely to account for much of this matter. Although the three diagnostic systems contain similar criteria for diagnosing insomnia, such as difficulties in initiating sleep or maintaining sleep, in addition to daytime dysfunctions or distress, the system of ICSD-II specifies an additional definition that insomnia occurs despite adequate opportunity and circumstances for sleep [51,52]. Of note, the DSM-V lists adequate sleep opportunity as one of the diagnostic criteria [53]. Nevertheless, all three instruments do not include this critical component, which may affect their accuracy as a screening tool for insomnia. Clinicians and researchers should evaluate whether individuals experience insomnia symptoms in situations where adequate sleep opportunity and circumstances for sleep are present before using these questionnaires.

A growing body of evidence suggests that the use of a case-control design can introduce spectrum bias [54] and thus produce inflated estimates of diagnostic performance compared with studies using a cohort of consecutive patients [49,55,56]. On average, the DOR was three times higher in studies evaluating these instruments in patients and a distinct healthy control group compared with those using a clinical population [55]. Concurs with this notion, the studies employing a case-control design to investigate the ISI had higher pooled sensitivity and specificity compared with those of studies using a cross-sectional design, probably because of the inclusion of individuals with an advanced target condition. Diagnostic case-control studies typically recruited severe cases that are relatively easy to diagnose and thus generate fewer false-negative test results compared with studies recruiting mild cases that are relatively difficult to diagnose.

A useful questionnaire for evaluating insomnia symptoms should consider diagnostic properties, sleep domains, and feasibility [57]. Regarding the diagnostic properties, the ISI, AIS, and PSQI yielded high and comparable sensitivity, suggesting that the three scales are suitable for assessing the risk of insomnia. Concerning specificity, the PSQI produced the lowest value (75%) compared with those of the ISI and AIS, which may increase false-positive results and thus lead to excessive distress in individuals who responded to the questionnaire. In terms of

Table 3
Summary estimates of sensitivity, specificity, and diagnostic odds ratio.

Scales	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)	Pooled DOR (95% CI)
ISI	0.88 (0.79 to 0.93)	0.85 (0.68 to 0.94)	41.9 (8.77 to 200.3)
AIS	0.91 (0.87 to 0.93)	0.87 (0.68 to 0.95)	67.7 (23.4 to 196.1)
PSQI	0.93 (0.86 to 0.96)	0.75 (0.64 to 0.84)	39.4 (14.5 to 106.9)
P value ISI vs. AIS	0.40	0.77	0.54
P value ISI vs. PSQI	0.31	0.31	0.98
P value AIS vs. PSQI	0.92	0.92	0.55

ISI = Insomnia Severity Index; AIS = Athens Insomnia Scale; PSQI = Pittsburgh Sleep Quality Index; CI = confidence interval; DOR = diagnostic odds ratio.

Table 4
Metaregression and moderator analysis.

Variables	n	Coefficient	Sensitivity	*P* value	Coefficient	Specificity	*P* value
ISI							
Age	6	0.11		0.69	0.18		0.36
Female percentage	7	−0.003		0.88	−0.008		0.81
Study design				<0.001			0.04
Case-control	3		0.94			0.94	
Others[a]	4		0.79			0.74	
Types of insomnia				0.33			0.57
Primary	3		0.91			0.89	
Secondary	4		0.84			0.82	
Reference test				0.15			0.35
ICSD-II	3		0.81			0.77	
Others[b]	4		0.92			0.89	
PSQI							
Age	8	0.03		0.49	0.03		0.17
Female percentage	8	0.03		0.41	−0.02		0.53
Region of study				0.39			0.81
Western countries	5		0.97			0.78	
Eastern countries	3		0.91			0.73	
Study design				0.96			0.64
Case-control	4		0.95			0.78	
Cross-sectional	4		0.96			0.74	
Types of insomnia				0.64			0.74
Primary	3		0.94			0.80	
Secondary	4		0.96			0.77	
Reference test				0.76			0.008
DSM-IV	6		0.93			0.81	
ICSD-II and a combination of DSM-IV and ICSD-R.	2		0.89			0.54	

CI = confidence interval; ISI = Insomnia Severity Index; PSQI = Pittsburgh Sleep Quality Index. *Psychology graduate students, research assistant, sleep physician.
[a] Cross-sectional study and retrospective cohort study.
[b] Insomnia Diagnostic Interview (a combination of DSM-IV, DSM-V and ICSD2), Insomnia Interview Schedule (a combination of DSM-IV and ICSD), DSM-IV, and DSM-IV-TR.

sleep domains, the ISI and AIS both assess all three major insomnia symptoms (i.e., difficulties in initiating sleep, maintaining sleep, early morning awakening), whereas the PSQI does not directly assess these insomnia symptoms but rather evaluates a broad range of sleep domains affecting sleep quality. Concerning feasibility, a previous study suggested that the ISI and AIS having fewer questions (7 items and 8 items) and requiring shorter administration time are superior to the PSQI [57]. Although the three scales together yield comparable diagnostic properties, the ISI and AIS were observed to rank the highest on the estimates of diagnostic accuracy as well as sleep domains assessed and feasibility for use as assessment tools compared with the PSQI. Taken together, we suggested that clinicians and researchers can apply the ISI and AIS in clinical practice or research as screening instruments for insomnia or as insomnia-related treatment outcomes. The PSQI is suitable as a measurement of patient-reported outcomes.

4.1. Limitations

Several limitations are evident while interpreting the outcomes of this review. First, the use of various reference standards, such as DSM-IV, ICD-10, and ICSD-II for identifying insomnia can produce varied estimates of the diagnostic accuracy and thus cause a difficulty in data combing. Furthermore, risks of bias in most domains were judged to be high or uncertain due to insufficient details in reporting. For example, blinding and test reproducibility were not fully reported, which can potentially alter data reliability. Third, because the pooled sensitivity and specificity of the three scales had relatively wide 95% CIs, indicating that precisions of the pooled results were not optimal probably due to the fact that only a small number of studies were included in our metaanalysis. The reliability and generalizability of the results might therefore be limited. Nevertheless, this meta-analysis has several strengths. This review adopted a robust search strategy and included a large sample size, which enhanced the internal validity of this review. In addition, including studies from various geographic areas substantially increased the generalizability of our findings.

5. Conclusion

In conclusion, we present that the ISI, AIS and PSQI are useful instruments for insomnia screening. Although no substantial evidence indicates which instrument is ideal compared with others for assessing the risk of insomnia, we suggest that the ISI and AIS are probably stronger and appropriate instruments according to the comparisons of scales characteristics for diagnostic properties, sleep domains and feasibility. Healthcare providers and researchers should be informed by our findings and select the most appropriate instrument for insomnia screening. Moreover, in future studies, investigators should enrol a cohort of consecutive patients or consider randomly selecting patients and blinding the assessor from the results of the index test and reference standard to provide an accurate estimate of instruments for insomnia screening.

Disclosure statement

This was not an industry supported study. This work was supported by a grant from the Ministry of Science and Technology, Taiwan (MOST 103–2314-B-038-068-MY2).

The authors declare no potential conflicts of interest with respect to the authorship and publication of this article, and report no financial disclosures.

Author contributions

Conception and design: Hsiao-Yean Chiu and Pei-Shan Tsai; analysis and interpretation: Hsiao-Yean Chiu, Ling-Yin Chang, Yu-Jung Hsieh, and Pei-Shan Tsai; drafting of the manuscript for important intellectual content: Hsiao-Yean Chiu and Pei-Shan Tsai.

References

[1] S. Ancoli-Israel, T. Roth, Characteristics of insomnia in the United States: results of the 1991 National Sleep Foundation Survey. I. Sleep. 22 (Suppl. 2) (1999) S347–S353.
[2] National Institutes of Health, National Institutes of Health State of the Science Conference statement on Manifestations and Management of Chronic Insomnia in Adults, June 13–15, 2005, Sleep 28 (2005) 1049–1057.
[3] C.C. Kao, C.J. Huang, M.Y. Wang, P.S. Tsai, Insomnia: prevalence and its impact on excessive daytime sleepiness and psychological well-being in the adult Taiwanese population, Qual. Life Res. 17 (2008) 1073–1080.
[4] M.M. Ohayon, C.F. Reynolds 3rd., Epidemiological and clinical relevance of insomnia diagnosis algorithms according to the DSM-IV and the International Classification of Sleep Disorders (ICSD), Sleep Med. 10 (2009) 952–960.
[5] M. Daley, C.M. Morin, M. LeBlanc, J.P. Gregoire, J. Savard, L. Baillargeon, Insomnia and its relationship to health-care utilization, work absenteeism, productivity and accidents, Sleep Med. 10 (2009) 427–438.
[6] S.C. Bolge, J.F. Doan, H. Kannan, R.W. Baran, Association of insomnia with quality of life, work productivity, and activity impairment, Qual. Life Res. 18 (2009) 415–422.
[7] W. Szelenberger, S. Niemcewicz, Severity of insomnia correlates with cognitive impairment, Acta Neurobiol Exp (Wars) 60 (2000) 373.
[8] W.W. Ishak, K. Bagot, S. Thomas, N. Magakian, D. Bedwani, D. Larson, et al., Quality of life in patients suffering from insomnia, Innov Clin Neurosci. 9 (2012) 13–26.
[9] T. Roth, New developments for treating sleep disorders, J Clin Psychiatry 62 (2001) 3–4.
[10] S. Saddichha, Diagnosis and treatment of chronic insomnia, Ann Indian Acad Neurol. 13 (2010) 94–102.
[11] H.U. Wittchen, P. Krause, M. Hofler, D. Pittrow, S. Winter, B. Spiegel, et al., NISAS-2000: the "Nationwide Insomnia Screening and Awareness Study". Prevalence and interventions in primary care, Fortschr Med Orig. 119 (2001) 9–19.
[12] D. Leger, B. Poursain, An international survey of insomnia: under-recognition and under-treatment of a polysymptomatic condition, Curr. Med. Res. Opin. 21 (2005) 1785–1792.
[13] M. Littner, M. Hirshkowitz, M. Kramer, S. Kapen, W.M. Anderson, D. Bailey, et al., Practice parameters for using polysomnography to evaluate insomnia: rupdate, Sleep 26 (2003) 754–760.
[14] E. Schramm, F. Hohagen, U. Grasshoff, D. Riemann, G. Hajak, H.G. Weess, et al., Test-retest reliability and validity of the Structured Interview for Sleep Disorders According to DSM-III-R, Am. J. Psychiatry 150 (1993) 867–872.
[15] C.H. Bastien, A. Vallieres, C.M. Morin, Validation of the Insomnia Severity Index as an outcome measure for insomnia research, Sleep Med. 2 (2001) 297–307.
[16] C.R. Soldatos, D.G. Dikeos, T.J. Paparrigopoulos, Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria, J. Psychosom. Res. 48 (2000) 555–560.
[17] D.J. Buysse, S. Ancoli-Israel, J.D. Edinger, K.L. Lichstein, C.M. Morin, Recommendations for a standard research assessment of insomnia, Sleep 29 (2006 Sep) 1155–1173.
[18] H.A. Lomeli, I. Perez-Olmos, C. Talero-Gutierrez, C.B. Moreno, R. Gonzalez-Reyes, L. Palacios, et al., Sleep evaluation scales and questionnaires: a review, Actas espanolas de psiquiatria. 36 (2008) 50–59.
[19] S.M. Alsaadi, J.H. McAuley, J.M. Hush, D.J. Bartlett, N. Henschke, R.R. Grunstein, et al., Detecting insomnia in patients with low back pain: accuracy of four self-report sleep measures, BMC Musculoskelet. Disord. 14 (2013) 196.
[20] N. Panic, E. Leoncini, G. de Belvis, W. Ricciardi, S. Boccia, Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses, PLoS One 8 (2013), e83138.
[21] R. Fluss, D. Faraggi, B. Reiser, Estimation of the Youden Index and its associated cutoff point, Biom. J. 47 (2005) 458–472.
[22] P.F. Whiting, A.W. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, et al., QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, Ann. Intern. Med. 155 (2011) 529–536.
[23] J.B. Reitsma, A.S. Glas, A.W. Rutjes, R.J. Scholten, P.M. Bossuyt, A.H. Zwinderman, Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews, J. Clin. Epidemiol. 58 (2005) 982–990.
[24] A.S. Glas, J.G. Lijmer, M.H. Prins, G.J. Bonsel, P.M. Bossuyt, The diagnostic odds ratio: a single indicator of test performance, J. Clin. Epidemiol. 56 (2003) 1129–1135.
[25] S.D. Walter, Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data, Stat. Med. 21 (2002) 1237–1256.
[26] J.P. Higgins, S.G. Thompson, Quantifying heterogeneity in a meta-analysis, Stat. Med. 21 (2002) 1539–1558.
[27] J.J. Deeks, P. Macaskill, L. Irwig, The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed, J. Clin. Epidemiol. 58 (2005) 882–893.
[28] K.J. Brower, J.T. Arbedt, D.A. Conroy, Comparing and validating three sleep scales in alcohol-dependent individuals, Am. J. Addict. 21 (2012) 391.
[29] G. Curcio, D. Tempesta, S. Scarlata, C. Marzano, F. Moroni, P.M. Rossini, et al., Validity of the Italian version of the Pittsburgh Sleep Quality Index (PSQI), Neurol. Sci. 34 (2013) 511–519.
[30] D. Marques, A. Allen Gomes, A. Meiavia, A. Salgueiro, C. Carlos Ribeiro, J. Dischler, Reliability and initial validation of the Pittsburgh Sleep Quality Index, European Portuguese version: a preliminary study in a sample of higher education students, Sleep Med. 14 (2013), e140.
[31] D.J. Buysse, C.F. Reynolds 3rd, T.H. Monk, S.R. Berman, D.J. Kupfer, The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research, Psychiatry Res. 28 (1989) 193–213.
[32] O.O. Aloba, A.O. Adewuya, B.A. Ola, B.M. Mapayi, Validity of the Pittsburgh Sleep Quality Index (PSQI) among Nigerian university students, Sleep Med. 8 (2007) 266–270.
[33] J. Backhaus, K. Junghanns, A. Broocks, D. Riemann, F. Hohagen, Test-retest reliability and validity of the Pittsburgh Sleep Quality Index in primary insomnia, J. Psychosom. Res. 53 (2002) 737–740.
[34] Y. Doi, M. Minowa, M. Uchiyama, M. Okawa, K. Kim, K. Shibui, et al., Psychometric assessment of subjective sleep quality using the Japanese version of the Pittsburgh Sleep Quality Index (PSQI-J) in psychiatric disordered and control subjects, Psychiatry Res. 97 (2000) 165–172.
[35] N.L. Fictenberg, S.H. Putnam, N.R. Mann, R.D. Zafonte, A.E. Millard, Insomnia screening in postacute traumatic brain injury: utility and validity of the Pittsburgh Sleep Quality Index, Am J Phys Med Rehabil. 80 (2001) 339–345.
[36] C. Gagnon, L. Belanger, H. Ivers, C.M. Morin, Validation of the Insomnia Severity Index in primary care, J. Am. Board Fam. Med. 26 (2013) 701–710.
[37] H.S. Jeong, Y. Jeon, J. Ma, Y. Choi, S. Ban, S. Lee, et al., Validation of the Athens Insomnia Scale for screening insomnia in South Korean firefighters and rescue workers, Qual. Life Res. 24 (2015) 2391–2395.
[38] C.M. Morin, G. Belleville, L. Belanger, H. Ivers, The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response, Sleep 34 (2011) 601–608.
[39] I. Okajima, S. Nakajima, M. Kobayashi, Y. Inoue, Development and validation of the Japanese version of the Athens Insomnia Scale, Psychiatry Clin. Neurosci. 67 (2013) 420–425.
[40] M.H. Savard, J. Savard, S. Simard, H. Ivers, Empirical validation of the Insomnia Severity Index in cancer patients, Psychooncology 14 (2005) 429–441.
[41] C.A. Severson, W.H. Tsai, P.E. Ronksley, S.R. Pendharkar, Identification of insomnia in a sleep center population using electronic health data sources and the insomnia severity index, J. Clin. Sleep Med. 9 (2013) 655–660.
[42] C.R. Soldatos, D.G. Dikeos, T.J. Paparrigopoulos, The diagnostic validity of the Athens Insomnia Scale, J. Psychosom. Res. 55 (2003) 263–267.
[43] J.L. Sun, J.F. Chiou, C.C. Lin, Validation of the Taiwanese version of the Athens Insomnia Scale and assessment of insomnia in Taiwanese cancer patients, J. Pain Symptom Manag. 41 (2011) 904–914.
[44] P.S. Tsai, S.Y. Wang, M.Y. Wang, C.T. Su, T.T. Yang, C.J. Huang, et al., Psychometric evaluation of the Chinese version of the Pittsburgh Sleep Quality Index (CPSQI) in primary insomnia and control subjects, Qual. Life Res. 14 (2005) 1943–1952.
[45] J.I. Tzeng, Y.W. Fu, C.C. Lin, Validity and reliability of the Taiwanese version of the Pittsburgh Sleep Quality Index in cancer patients, Int. J. Nurs. Stud. 49 (2012) 102–108.
[46] C.M. Yang, S.C. Hsu, S.C. Lin, Y.Y. Chiu, I.Y.M. Chen, Reliability and validity of the Chinese version of insomnia severity index (Chinese), Arch Clini Psychol. 4 (2009) 95–104.
[47] J. Farrahi Moghaddam, N. Nakhaee, V. Sheibani, B. Garrusi, A. Amirkafi, Reliability and validity of the Persian version of the Pittsburgh Sleep Quality Index (PSQI-P), Sleep Breath. 16 (2012) 79–82.
[48] Y.W. Cho, M.L. Song, C.M. Morin, Validation of a Korean version of the insomnia severity index, J Clin Neurol. 10 (2014) 210–215.
[49] A.W. Rutjes, J.B. Reitsma, J.P. Vandenbroucke, A.S. Glas, P.M. Bossuyt, Case-control and two-gate designs in diagnostic accuracy studies, Clin. Chem. 51 (2005) 1335–1341.
[50] R.C. Kessler, C. Coulouvrat, G. Hajak, M.D. Lakoma, T. Roth, N. Sampson, et al., Reliability and validity of the brief insomnia questionnaire in the America insomnia survey, Sleep 33 (2010) 1539–1549.
[51] M.J. Thorpy, Classification of sleep disorders, Neurotherapeutics 9 (2012) 687–701.
[52] K.F. Chung, W.F. Yeung, F.Y. Ho, K.P. Yung, Y.M. Yu, C.W. Kwok, Cross-cultural and comparative epidemiology of insomnia: the Diagnostic and statistical manual (DSM), International classification of diseases (ICD) and International classification of sleep disorders (ICSD), Sleep Med. 16 (2015) 477–482.
[53] Association AP, DSM-5 Classification, American Psychiatric Association, 2016.
[54] D.F. Ransohoff, A.R. Feinstein, Problems of spectrum and bias in evaluating the efficacy of diagnostic tests, N. Engl. J. Med. 299 (1978) 926–930.
[55] J.G. Lijmer, B.W. Mol, S. Heisterkamp, G.J. Bonsel, M.H. Prins, J.H. van der Meulen, et al., Empirical evidence of design-related bias in studies of diagnostic tests, JAMA 282 (1999) 1061–1066.
[56] P.F. Whiting, A.W. Rutjes, M.E. Westwood, S. Mallett, A systematic review classifies sources of bias and variation in diagnostic test accuracy studies, J. Clin. Epidemiol. 66 (2013) 1093–1104.
[57] G.A. Wells, T. Li, J.R. Kirwan, J. Peterson, D. Aletaha, M. Boers, et al., Assessing quality of sleep in patients with rheumatoid arthritis, J. Rheumatol. 36 (2009) 2077–2086.