Journal of Psychosomatic Research 78 (2015) 34–38

Contents lists available at ScienceDirect

Journal of Psychosomatic Research

Comparison of scoring methods for the Brief Insomnia Questionnaire in a general population sample

Ka-Fai Chung [a,*], Wing-Fai Yeung [b], Fiona Yan-Yee Ho [c], Lai-Ming Ho [d], Kam-Ping Yung [e], Yee-Man Yu [a], Chi-Wa Kwok [a]

[a] Department of Psychiatry, University of Hong Kong, Hong Kong, China
[b] School of Chinese Medicine, University of Hong Kong, Hong Kong, China
[c] Department of Psychology, University of Hong Kong, Hong Kong, China
[d] School of Public Health, University of Hong Kong, Hong Kong, China
[e] Department of Psychology, Chinese University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:
Received 20 May 2014
Received in revised form 30 October 2014
Accepted 14 November 2014

Keywords:
Diagnosis
DSM-IV-TR
DSM-5
ICD-10
ICSD-2
ICSD-3
Insomnia

ABSTRACT

Objective: The Brief Insomnia Questionnaire (BIQ) is a lay-administered, structured interview to derive insomnia disorders according to the Diagnostic and Statistical Manual, Fourth Edition, Text Revision (DSM-IV-TR), International Classification of Diseases, Tenth Edition (ICD-10) and research diagnostic criteria/International Classification of Sleep Disorders, Second Edition (RDC/ICSD-2). The concordance between diagnoses derived from the BIQ and clinical interviews was only moderate and the prevalence estimates based on the BIQ were significantly different from estimates based on clinical interviews. We hypothesized that a modification of the scoring algorithm closer to the diagnostic criteria would improve the performance of the BIQ.
Methods: Probability subsample of population-based epidemiological survey respondents (n = 2011) completed clinical reappraisal (n = 176) interviews. We compared the modified scoring with the original scoring in sensitivity, specificity, positive and negative predictive values, areas under the characteristic curve, and Cohen's kappa to detect DSM-IV-TR, ICD-10 and RDC/ICSD-2 insomnia diagnoses by the BIQ against clinical interviews.
Result: The diagnostic accuracy was improved with the modified scoring. The areas under the receiver operating characteristic curve for the DSM-IV-TR, ICD-10, RDC/ICSD-2 and any of the insomnia diagnoses ranged from 0.76 to 0.87. Using the modified scoring, there was no significant difference between prevalence estimates based on the BIQ classification and clinical interviews.
Conclusions: The BIQ with modified scoring enhanced case detection and produced more accurate prevalence estimates of DSM-IV-TR, ICD-10 and RDC/ICSD-2 insomnia disorders. With scoring algorithms now extended to DSM-5 and ICSD-3 diagnoses, the BIQ should be more widely used in clinical and research settings.

Introduction

Insomnia is the most common sleep complaint and it is associated with daytime impairments and health risks. Epidemiological studies have reported the prevalence of insomnia ranging from 6% to 48% [1]. In Hong Kong, the prevalence of insomnia in the general population was estimated to be 11.9% [2] and 39.4% in another study [3]. Estimates of insomnia prevalence vary widely due to the use of different diagnostic criteria, questionnaires and interview methods. The Sleep-EVAL interview was developed by Ohayon et al. [4] in the late 1990s to derive sleep disorder diagnoses according to the Diagnostic and Statistical Manual of Mental Disorder, Fourth Edition (DSM-IV) and International Classification of Sleep Disorders (ICSD), but it is not generally available to the insomnia research community. Recently, a standardized lay-administered instrument, the Brief Insomnia Questionnaire (BIQ), was developed for use in the America Insomnia Survey [5]. The BIQ covers the diagnostic criteria of insomnia disorder according to the DSM-IV, Text Revision (DSM-IV-TR) [6], the International Classification of Diseases, Tenth Edition (ICD-10) [7] and research diagnostic criteria/ICSD, Second Edition (RDC/ICSD-2) [8,9]. Following the proposed scoring algorithms [5], insomnia disorder diagnoses by the DSM-IV-TR, ICD-10 and RDC/ICSD-2 can be obtained; making the BIQ useful for comparative epidemiological studies and for clinicians and researchers who may have different preference of diagnostic system.

The original validation study showed that diagnoses derived from the BIQ had moderate concordance with diagnoses based on semi-structured clinical reappraisal interviews, with sensitivity from 41.7% to 67.6% and specificity from 94.4% to 98.5%. There was also a significant difference between BIQ and clinical-interview estimates of the prevalence of insomnia disorder, compromising the BIQ as an epidemiological

* Corresponding author at: Department of Psychiatry, The University of Hong Kong, Pokfulam Road, Hong Kong, China. Tel.: +852 22554487; fax: +852 28551345.
E-mail address: kfchung@hkucc.hku.hk (K.-F. Chung).

http://dx.doi.org/10.1016/j.jpsychores.2014.11.015
0022-3999/

tool. The discrepancy in prevalence estimates was also found in our validation study of the Hong Kong Chinese version of the BIQ (HK-BIQ) [10]. Compared with estimates by reappraisal interviews, the HK-BIQ was accurate in estimating the prevalence of RDC/ICSD-2 insomnia disorder, but under-estimated the DSM-IV-TR and over-estimated the ICD-10 insomnia diagnoses.

To address this issue, we reviewed the scoring method of the BIQ and modified the scoring to approximate with the diagnostic criteria. The DSM-IV-TR criterion B requires an individual's sleep disturbance causing clinically significant distress or impairment in social, occupational, or other important areas of functioning. A similar wording is used in the ICD-10 criterion D. We modified the scoring to allow the criterion be met if any one of the BIQ daytime impairment items is positive, instead of two in the original scoring algorithms. In addition, the ICD-10 criterion C requires that there is preoccupation with the sleeplessness and excessive concern over its consequences at night and during the day. We modified the scoring such that the criterion is met when both the BIQ excessive concern item and the worries or distress item are present, instead of either one of the items is present in the original scoring algorithm. The scoring algorithm for RDC/ICSD-2 insomnia disorder was found to be compatible with the diagnostic criteria; hence, we had not made any adjustment. The scoring algorithms that have been modified are presented in Appendix A.

We believe that the closer the scoring algorithms are in line with the diagnostic criteria, the greater will be the likelihood of agreement between the BIQ and clinical diagnosis and closer will be the prevalence estimates. The recent launch of the DSM-5 [11] and ICSD-3 [12] might have hampered the need of the old systems; however, comparative epidemiological studies of the change after using the new systems are important, especially for the ICSD, which has mainly revised the duration criterion from 1 month in RDC/ICSD-2 to 3 months in ICSD-3. Using the same sample and statistical approach as in the original HK-BIQ study [10], we compared the diagnostic accuracy between the original and modified scoring method, with a view that the modified scoring method will provide more accurate case detection and prevalence estimates.

Method

Sample

The sample and study procedure were reported in detail in a previous study [10]. The study population consisted of Hong Kong residents older than 18 years and able to communicate in Cantonese or Mandarin Chinese languages. The randomization process was divided into 2 parts: randomization of telephone numbers and selection of respondents in households. Telephone numbers in Hong Kong are listed in telephone directories automatically unless the customers request their numbers be withheld. We selected telephone numbers randomly from computerized residential telephone directories with no stratification applied and generated some unlisted numbers by adding and subtracting 1 and 2 from the selected numbers [13]. Duplicate numbers were screened out. Within each household, respondents were randomly selected by asking to speak to the person who was going to celebrate a birthday next. This technique is commonly used to overcome respondent selection bias associated with administering the survey to the household member most likely to answer the phone. A recent review detected no significance differences in demographic distribution between "next birthday" and true probability samples [14]. Verbal consent was obtained from all participants and all procedures used in this study were reviewed and approved by the local institutional review board.

Procedure

A fully-structured lay-administered telephone interview was conducted by the Public Opinion Programme, The University of Hong Kong. We successfully interviewed 2011 respondents from July 24 to December 6, 2012. The overall response rate was 64.3%. There were 1019 refusals at household or respondent-levels and 97 partial responses. The first section included an introduction and verbal consent, followed by the HK-BIQ, then sociodemographics, including age, gender, occupation and level of education. The last section consisted of verbal consent to another telephone interview on their sleep problem. In most cases, the telephone interview could be completed within 15 min. In line with the American BIQ validation study [5], we randomly selected participants but oversampled BIQ positives for clinical reappraisal interviews. The clinical reappraisal subsample consisted of 73 cases, 51 subthreshold cases and 52 non-cases, which allowed a Cohen's kappa (κ) of 0.7 with 2-sided 95% confidence interval (CI) of 0.1 [15]. The telephone-based clinical reappraisal interviews were conducted in a blinded manner 2–14 days after the first interview by 2 senior authors (KC and WY). The re-interviews were timed to minimize memory effects and true changes in insomnia status and to provide some flexibility in contacting the respondents. No respondents were paid for participation.

Measures

The translation of the BIQ into Chinese was conducted according to the World Health Organization guidelines [16], with steps including forward translation, expert review, back-translation, expert review, pre-testing, and final version. The expert panel consists of experienced clinicians and researchers in sleep disorders. The clinical reappraisal was conducted using a standardized semi-structured questionnaire, developed specifically for BIQ validation [5]. It includes DSM-IV-TR, ICD-10 and RDC/ICSD-2 symptom checklists, with rating categories of definite, probable, possible and no for each symptom and classification as case or non-case for each diagnostic system. The κ values between KC and WY for all diagnostic categories were 1.0 based on 20 audiotaped interviews.

Data analysis

All statistical analysis was done by STATA 10.0. The clinical reappraisal subsample was first weighted to adjust for the oversampling of BIQ positives. Validity of the BIQ was assessed by the concordance between diagnoses based on the BIQ and diagnoses based on clinical reappraisal. The validity and reliability at both aggregate and individual levels were tested. At the aggregate level, we compared prevalence estimates based on the BIQ and the clinical reappraisal by the McNemar χ^2 tests. Individual-level diagnostic concordance was evaluated using 2 descriptive measures, the area under the receiver operating characteristic curve (AUC) [17] and κ. AUC was calculated as $(TP / P + TN / N) / 2$ while κ as $[TP - FP - P'(1 - 2N)] / [P - P'(1 - 2N)]$, where TP = true positive, FP = false positive, TN = true negative, FN = false negative, $P = TP + FN$, $P' = TP + FP$, and $N = FP + TN$ [18]. Although AUC and κ are related concepts, they may not always change in the same direction. We also reported sensitivity (SN), specificity (SP), positive predictive value (PPV) and negative predictive value (NPV). The odd ratio, which is equal to $[SN \times SP] / [(1 - SN) \times (1 - SP)]$ was also used to assess concordance between diagnoses based on the BIQ and the clinical reappraisal. We used net reclassification index (NRI) to examine the change in diagnostic accuracy after using the modified scoring algorithm. Overall NRI was calculated using the formula: $[Pr(up|event) - Pr(down|event)] + [Pr(down|nonevent) - Pr(up|nonevent)]$, with higher values indicating greater improvement [19]. To determine whether BIQ symptom-level data improved the prediction of clinical diagnoses, a series of stepwise logistic-regression equations, in which clinical diagnoses were treated as dichotomous outcomes and BIQ symptom variables were included along with BIQ diagnoses as predictors. We compared the AUC based on dichotomous BIQ classification with that of continuous predicted