Sleep Medicine 35 (2017) 35–40

Contents lists available at ScienceDirect

Sleep Medicine

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

Original Article

Are the Insomnia Severity Index and Pittsburgh Sleep Quality Index valid outcome measures for Cognitive Behavioral Therapy for Insomnia? Inquiry from the perspective of response shifts and longitudinal measurement invariance in their Chinese versions

Po-Yi Chen [a], Ya-Wen Jan [b, c], Chien-Ming Yang [c, d, *]

[a] Department of Psychology University of Kansas, USA
[b] Sleep Research Center, Taipei Medical University Hospital, Taipei, Taiwan
[c] Department of Psychology, National Chengchi University, Taipei, Taiwan
[d] The Research Center for Mind, Brain, and Learning, National Chengchi University, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 11 January 2017
Received in revised form
20 March 2017
Accepted 6 April 2017
Available online 30 April 2017

Keywords:
Cognitive Behavioral Therapy for Insomnia
CBT-I
Insomnia Severity Index
Pittsburgh Sleep Quality Index
Measurement invariance
Response shift

ABSTRACT

Objective: The purpose of this study was to examine whether the Insomnia Severity Index (ISI) and Pittsburgh Sleep Quality Index (PSQI) are valid outcome measures for Cognitive Behavioral Therapy for Insomnia (CBT-I). Specifically, we tested whether the factorial parameters of the ISI and the PSQI could remain invariant against CBT-I, which is a prerequisite to using their change scores as an unbiased measure of the treatment outcome of CBT-I.
Methods: A clinical data set including scores on the Chinese versions of the ISI and the PSQI obtained from 114 insomnia patients prior to and after a 6-week CBT-I program in Taiwan was analyzed. A series of measurement invariance (MI) tests were conducted to compare the factorial parameters of the ISI and the PSQI before and after the CBT-I treatment program.
Results: Most factorial parameters of the ISI remained invariant after CBT-I. However, the factorial model of the PSQI changed after CBT-I treatment. An extra loading with three residual correlations was added into the factorial model after treatment.
Conclusions: The partial strong invariance of the ISI supports that it is a valid outcome measure for CBT-I. In contrast, various changes in the factor model of the PSQI indicate that it may not be an appropriate outcome measure for CBT-I. Some possible causes for the changes of the constructs of the PSQI following CBT-I are discussed.

1. Introduction

Insomnia is defined as the subjective complaint of dissatisfaction with sleep quantity and/or quality [1]. Thus, a self-rating measure for the severity of insomnia is important in studies of insomnia treatments. Researchers often use the difference scores on self-rating scales assessed before and after treatment as an index for patients' improvement associated with the treatment [2]. Although the use of change scores for treatment effect is a common practice, potential threats that might confound the interpretations of the scores are often neglected. From the perspective of psychometrics, for example, longitudinal measurement invariance (MI) against the treatment should be a prerequisite for using the change on a scale as a measure for treatment outcome. A lack of MI can bias the interpretation of the change scores and therefore threaten the validity of the results [3].

1.1. Importance of MI in evaluating the effects of psychological interventions

MI is an important element of psychological tests. It concerns whether the target construct is measured in the same way across occasions [4,5]. In longitudinal studies, testing MI is similar to examining whether researchers use the same ruler to measure a target across time points [6], where the “scales of the ruler” are the

* Corresponding author. Department of Psychology, National Chengchi University, 64, Sec. 2, Chih-Nan Rd., Taipei, 116, Taiwan. Fax: +886 2 29390644.
E-mail address: yangcm@nccu.edu.tw (C.-M. Yang).

http://dx.doi.org/10.1016/j.sleep.2017.04.003
1389-9457/

factorial parameters like factor loadings or intercepts. The longitudinal change scores of a self-report measure can be meaningfully interpreted only when the "scale" of the ruler is identical across time points. Otherwise, researchers will have difficulty differentiating the "true change" of the targets (eg, decrease in severity of insomnia due to Cognitive Behavioral Therapy for Insomnia [CBT-I]) from simple shifts in the scale of the ruler.

In the scenario of CBT-I, the treatment might not only alleviate the symptoms of insomnia; it could also change patients' attitudes and concepts of sleep and insomnia. Given that previous studies have shown that the cognitive changes caused by psychological interventions might change the factorial parameters underlying questionnaires, it is reasonable to suspect that some questionnaires that researchers use in CBT-I could also be affected. These changes could hinder researchers in accurately estimating the treatment efficacy of CBT-I. For example, if CBT-I will strengthen an item's relation with the underlying construct of insomnia in a patient, then the factor loading of this item could also be increased, because loadings are usually considered as links between observable indicators to latent constructs. Given this situation, the observed change score on this item for insomnia severity will, on average, underestimate the "real" treatment efficacy, for the decrease in subjective ratings on the targeted latent construct will be offset by the increase of factor loadings [7–10]. More detailed explanations of the influences of noninvariant factorial parameters on the validity of change scores can be found in Appendix (online supplementary materials).

From the perspective of psychometrics, checking whether factorial parameters are invariant across times is exactly the issue addressed by MI tests in confirmatory factor analysis (CFA). The four most common MI tests in CFA are configural, weak, strong, and strict invariance tests. These four tests focus on the factor structure, loadings, intercepts, and residual variances respectively, and are usually tested in sequence. According to the literature [7], strong invariance is a key property to ensure the longitudinal comparability of a self-report measure.

1.2. Response shifts caused by cognitive behavioral therapy (CBT) and their corresponding clinical interpretations in a CFA framework

In contrast to the invariant properties mentioned above, the changes (ie, noninvariance) in factor loadings and intercepts can be associated with a psychological phenomenon called response shifts [11–13]. A response shift can be defined as a change in the meaning of one's self-report on the target construct due to the following: 1) changes in the internal scale that one uses for self-evaluation (ie, recalibration); 2) rearrangements in the order of importance of the items (components) that one uses to compose the target construct (ie, reprioritization); or 3) conceptual changes in the way that one defines the target construct (ie, reconceptualization) [14].

Various methods have been developed to detect response shifts. Oort proposed a procedure to test response shifts with invariance tests in CFA (ie, to identify the non-invariance parts of a factorial model) [12]. Fokkema et al. extended Oort's works to the scenario of CBT for depression and offered possible clinical interpretations of different kinds of noninvariance [11]. They proposed that changes in a questionnaire's factor structure across times (ie, failure to pass the configural invariance test) can be considered as evidence of reconceptualization, because it means that patients used different items to define the underlying construct after the treatment of CBT. Second, if the factor structure underlying a questionnaire is invariant after treatment but the loadings of some items become higher (ie, failure to pass the weak invariance test), then it represents a reprioritization because these items have become more indicative for the patients. Third, the changes in items' intercepts (failure to pass the strong invariance test) represent uniform recalibration (change in the initial point of self-evaluation). An increase in an item's intercept might indicate that patients have become more sensitive to the symptom depicted by the item after CBT treatment. Fourth, changes in residual variance can be considered nonuniform recalibration. Table 1 presents a summary of Oort's definition of response shifts (in a CFA framework), corresponding MI tests, and possible clinical interpretations proposed by Fokemma et al.

1.3. Response shifts that have been found in psychological interventions

Considering that the aims of psychological treatments such as CBT usually involve reshaping patients' cognitions, it is reasonable that some parts of the factorial model will be changed after treatment [3,11,15–17]. For example, researchers have recently found evidence indicating that the Beck Depression Inventory (BDI) is not invariant against CBT for depression [11]. Specifically, it was found that most of the intercepts and two-factor loadings of the BDI changed after treatment. This phenomenon not only reflects the response shifts caused by CBT but also indicates that the BDI failed to pass the strong invariance test. Given these results, the authors concluded that the scores on the BDI obtained before and after CBT might not be comparable to each other. Wu also found similar results [15] and concluded that, due to the confounding caused by response shifts (ie, measurement noninvariance), using the BDI as a measure for psychological treatment outcome may entail bias.

1.4. Possible response shifts on the ISI and PSQI caused by CBT-I

Several studies have demonstrated the potential influence of response shifts (noninvariance) on self-report measures after psychological interventions. Consequently, researchers in different areas have begun to examine the MI of their own measures with MI tests after psychological interventions [11,15–17]. Among the treatments for insomnia, CBT-I has been demonstrated to be effective and is recommended as a first-line treatment for insomnia [18]. CBT-I, like many other psychological interventions, also contains the element of reshaping patients' cognitions and could therefore cause response shifts (noninvariance). As far as we know, no studies to date have addressed the MI of the outcome measures in treatment studies of CBT-I.

Among the rating scales commonly used in insomnia studies, the Insomnia Severity Index (ISI) and the Pittsburgh Sleep Quality Index (PSQI) are the recommended measures for global sleep and insomnia symptoms in a standard research assessment protocol for insomnia [19–21]. The ISI is one of the most common outcome measures in CBT-I studies [22–25]. The PSQI is also widely used as an outcome measure in CBT-I studies [2,25,26]. For example, in a recent systematic review and meta-analysis of CBT-I for chronic insomnia, the PSQI is one of the two questionnaires identified to be used consistently enough for meta-analysis at the posttreatment time point [25]. Another recent meta-analysis of the treatment effect of group CBT-I, PSQI was used in four of the seven studies included in the study [26]. As a result, in the current study, we examined the MI properties of the ISI and the PSQI against CBT-I.

1.5. Research hypotheses

Most of the items on the ISI directly focus on patients' subjective feelings about their insomnia symptoms [21]. Furthermore, the three-factor model of ISI proposed by Bastien et al. [22] has been successfully replicated in clinical, nonclinical, and cross-cultural studies [27,28]. In contrast, the PSQI was developed to measure