Temporal Stability of the Ford Insomnia Response to Stress Test (FIRST)

Denise C. Jarrin, PhD¹,²; Ivy Y. Chen, BS¹,²; Hans Ivers, PhD¹,²; Christopher L. Drake, PhD³; Charles M. Morin, PhD¹,²

¹École de Psychologie, Université Laval, Québec City, Québec, Canada;
²Centre d’étude des Troubles du Sommeil, Centre de Recherche de l’Institut Universitaire en Santé Mentale de Québec, Québec City, Québec, Canada;
³Henry Ford Hospital, Detroit, MI

Acknowledgement / Funding: This study was supported by the Canadian Institutes of Health Research (MOP42504 and B0512201) and the Fonds de recherche du Québec – Santé (32207). Dr. Jarrin is funded by the Fonds de recherche du Québec – Santé (32207).

Correspondence: Charles M. Morin, PhD, 2325, rue des Bibliothèques; École de Psychologie, Université Laval, Québec, Canada, G1V 0A6; Email: cmorin@psy.ulaval.ca
Abstract

**Study Objectives:** The Ford Insomnia Response to Stress Test (FIRST) is a self-report tool that measures sleep reactivity (i.e., vulnerability to experience situational insomnia under stressful conditions). Sleep reactivity has been termed a “trait-like” vulnerability; however, evidence of its long-term stability is lacking. The main objective of the current psychometric study was to investigate the temporal stability of the FIRST over two 6-mo intervals in a population-based sample of adults with and without insomnia. The temporal stability of the FIRST was also compared with the temporal stability of other scales associated with insomnia (trait-anxiety, arousability).

**Methods:** Participants included 1,122 adults (mean age = 49.9 y, standard deviation = 14.8; 38.8% male) presenting with an insomnia syndrome (n = 159), insomnia symptoms (n = 152), or good sleep (n = 811). Participants completed the FIRST, the State-Trait Anxiety Inventory (trait-anxiety), and the Arousal Predisposition Scale (arousability) on three different occasions: baseline and at 6- and 12-mo follow-up. Intraclass correlation coefficients (ICCs) were computed for all scales (baseline to 6 mo and 6 to 12 mo).

**Results:** The FIRST yielded strong temporal stability from baseline to 6 mo among those with insomnia syndrome (ICC = 0.81), symptoms (ICC = 0.78), and good sleep (ICC = 0.81). Similar results were observed for 6 to 12 mo among those with insomnia syndrome (ICC = 0.74), insomnia symptoms (ICC = 0.82), and good sleep (ICC = 0.84). The stability of the FIRST was not comparable with the stability of trait-anxiety, but was somewhat comparable with the stability of arousability.

**Conclusions:** Overall, the FIRST is a temporally reliable stable scale over 6-mo intervals. Future research is needed to corroborate the stability and trait-like measures of sleep reactivity with physiological, behavioural and personality measures.

**Keywords:** insomnia, psychometrics, reliability, sleep reactivity, temporal stability
BRIEF SUMMARY

Current Knowledge/Study Rationale: The use of the Ford Insomnia Response to Stress Test (FIRST) has grown in the field; however, data on its temporal stability are scarce and based on relatively short-term test-retest intervals in small samples. The current psychometric study aimed to investigate the temporal stability of the FIRST over two 6-mo intervals in a population-based sample of adults with and without insomnia.

Study Impact: Such information is important when assessing the psychometric properties of the FIRST (i.e. high FIRST scores). This is the first step in its future use to identify and predict long-term risk and outcomes in individuals with sleep reactivity.
INTRODUCTION

Predisposing factors are individual traits or characteristics that exist prior to the onset of insomnia, and under precipitating events (e.g., stress), can increase one’s susceptibility to experience insomnia. Several personality and psychological characteristics are recognized as putative predisposing traits of insomnia, including a ruminative personality, high neuroticism, high depression, anxiety introversion, and arousability. Another predisposing factor of insomnia, which has gained attention in the literature, is a traitlike vulnerability to stress-related insomnia termed sleep reactivity.

Sleep reactivity is proposed to be a catalyst of insomnia as a result of exposed and/or perceived stressful environments in vulnerable individuals. The Ford Insomnia Response to Stress Test (FIRST), a self-report tool, was developed to measure and identify individuals with sleep reactivity. Participants rate how likely the experience of common stressful scenarios lead to the disruption in their sleep. Under experimentally induced (e.g., following caffeine consumption) or naturally occurring stressful settings (e.g., work stress), good sleepers with high FIRST scores experience more sleep difficulties and exhibit elevated physiological stress reactivity than those with low FIRST scores.

Substantial evidence demonstrates reliability, validity, and utility for the FIRST. The items on the FIRST show good levels of discriminant validity and high levels of internal consistency in both males and females, in young and middle-aged adults, and in clinical and population-based samples. Longitudinal studies also demonstrate the predictive validity of the FIRST. High FIRST scores increase the risk of new-onset insomnia and predict the chronicity of insomnia in good sleepers up to 3 y later.

Additionally, several cross-sectional studies have identified strong associations between FIRST scores with traits and/or general dimensions of personality commonly exhibited among insomnia patients, hence supporting its convergent validity. For
example, high FIRST scores are positively correlated with presleep somatic and cognitive arousal, neuroticism, and anxiety trait measures. More recently, clinically relevant cutoff values for the FIRST (i.e., > 16 and > 18) have been identified and are being used as a preventative tool to identify at-risk individuals for acute and chronic insomnia, respectively.

Sleep reactivity, as measured by the FIRST, has been termed a trait-like vulnerability. Based on the trait theory, traits differ across people, affect behavior, and importantly, are stable patterns (in behavior, cognitions, and emotions) of habit over time. Though current evidence indicates sleep reactivity indeed differs across individuals (i.e., with varying levels of vulnerability in experiencing situational insomnia), negatively affects sleep under stressful conditions (e.g., longer sleep onset latency, nocturnal arousals), and is associated with maladaptive behaviors (e.g., sleep-interfering behaviors), cognitions (e.g., rumination, dysfunctional sleep beliefs), and emotions (e.g., emotion-oriented coping), evidence of its temporal stability is lacking (i.e., whether the FIRST is reliably stable over time?).

Although the attention on and use of the FIRST has grown in the field, data on the temporal stability of the FIRST are based on relatively short-term test-retest intervals in small samples. In its original publication, the FIRST test-retest reliability was 0.92 after 2 w in a sample of 10 participants. More recently, in a sample of 45 participants, the test-retest reliability of a French version of the FIRST was 0.75 after 3 w. Given the sparse test-retest data of the FIRST and the importance of such psychometric data given its wide-ranging use in the field, the main objective of the current study was to investigate the temporal stability of the FIRST over two 6-month intervals (i.e., baseline to 6 mo and 6 to 12 mo) in a population-based sample of adults with and without insomnia. The temporal stability of the FIRST was also compared with the stability of other validated scales that assess traits well known to correlate with insomnia (trait-
It was posited that the FIRST would show high and similar temporal stability as other recognized traits.

METHODS

Study Context

Data from this study were derived from a larger population-based epidemiological study conducted in Québec, Canada. This study context and sample selection has been previously described. Briefly, the goal of the epidemiological study was to document the natural history and treatment of insomnia. The study began with a telephone survey using random digit dialing and the Kish method. At the conclusion of the telephone interview, participants were asked if they wanted to take part in the longitudinal phase of the study. The baseline evaluation was conducted 1 mo after the telephone interview and the remaining evaluations were conducted at baseline, 6 mo, and 12 mo after the first postal evaluation. Several additional evaluations were conducted but only data from the baseline evaluation and the following two postal evaluations (1-y period) were used in the current study. Ethical approval was obtained by Université Laval and all participants gave written informed consent.

Participants

Following postal evaluation, participants were classified into one of three sleep status groups: insomnia syndrome, insomnia symptoms, and good sleep. Given the waxing-waning feature of insomnia, only participants with consistent sleep patterns across all three assessments (i.e., baseline, 6- and 12-mo follow-up) were included in the current study. There were 159 participants with insomnia syndrome (mean age = 55.1 y, standard deviation (SD) = 13.7; 24.5% male), 152 with insomnia symptoms (mean age = 51.3 y, SD = 14.2; 30.3% male), and 811 with good sleep (mean age = 48.6 y, SD = 14.9; 43.2% male).
Sleep Status Groups

The classification for the sleep status groups were based on an algorithm that used a combination of diagnostic criteria for insomnia disorder from the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV-TR)\(^24\) and the International Classification of Diseases, 10th Edition (ICD-10),\(^25\) as well as the reported use of sleep-promoting medications or products. Responses from the Insomnia Severity Index (ISI),\(^{26,27}\) the Pittsburgh Sleep Quality Index (PSQI),\(^{28}\) and from questions on the use of sleep medication were used to determine the presence or absence of each criterion (described in the next paragraphs).

**Insomnia Syndrome**

Participants with insomnia syndrome met all diagnostic criteria for insomnia. These included being dissatisfied with their sleep patterns and presented symptoms of initial (sleep onset latency > 30 min), middle (wake time after sleep onset > 30 min), or late (morning awakening > 30 min before desired wake time) insomnia at least 3 nights per week for at least 1 mo. Another criterion included substantial psychological distress or impairments in daytime functioning related to sleep difficulties. Participants who used prescribed medication as a sleep-promoting agent more than 3 nights per week were also classified into this group.

**Insomnia Symptoms**

Participants with insomnia symptoms presented symptoms of initial, middle, or late insomnia at least 3 nights per week without meeting all the diagnostic criteria for an insomnia syndrome (i.e., did not report psychological distress). Participants who were dissatisfied with their sleep quality but did not present symptoms of initial, middle, or late insomnia and who used prescribed medication or over-the-counter sleep-promoting products at least 1 night but less than 3 nights per week were classified into this group.
Good Sleepers

Good sleepers were satisfied with their sleep quality, did not present symptoms of initial, middle, or late insomnia, and did not use prescribed or over-the-counter medication to promote sleep.

Measures

A battery of validated self-report questionnaires were sent to participants, covering sleep, mood, life events, and health-related quality of life domains; however, only independent and dependent variables and covariates used in the current study will be described in detail.

Insomnia Severity Index

The ISI\textsuperscript{26,27} is a seven-item, self-rated scale used to assess perceived insomnia severity based on items related to sleep problems, sleep satisfaction, interference of sleep difficulties with daytime functioning, noticeability of sleep problems, and distress caused by sleep difficulties over the past month. Each item, rated on a five-point Likert scale, is summed to yield a total score (range: 0-28), with higher scores indicative of severe insomnia. The ISI has adequate psychometric properties and is sensitive to treatment responses.\textsuperscript{27}

The Pittsburgh Sleep Quality Index

The PSQI\textsuperscript{28} is a 19-item, self-rated scale used to assess sleep quality and sleep patterns over the past month. Items are used to derive seven clinically based subscales (sleep quality, sleep latency, sleep duration, use of sleeping medication, habitual sleep efficiency, sleep disturbances, and daytime dysfunction). Subscales are summed to obtain the Global PSQI score (range: 0-21), with higher scores indicative of considerable sleep disturbance and poor sleep quality. Psychometric properties of the PSQI are well recognized.\textsuperscript{28}
**Ford Insomnia Response to Stress Test**

The FIRST\(^7\) is a nine-item scale used to assess an individual's likelihood of experiencing sleep difficulties in response to common stressful situations. Each item is self-rated on a four-point Likert scale and summed to yield a total score (range: 9-36); higher scores indicate higher levels of sleep reactivity.

**State-Trait Anxiety Inventory, Trait Anxiety Scale (STAI-T)**

The trait part of the State-Trait Anxiety Inventory, Trait Anxiety Scale (STAI-T)\(^{29}\) is a 20-item scale used to assess anxiety as a personality trait. Participants are asked to score how they relate to the statements (in general) on a four-point Likert scale. Scores are summed, with higher scores indicative of a greater level of anxiety (range: 20-80). The psychometric properties of the STAI-T\(^{30}\) in the French version are well documented.\(^{31}\)

**The Arousal Predisposition Scale**

The Arousal Predisposition Scale (APS)\(^{32}\) is a 12-item self-report scale used to assess arousability. Participants report the frequency of experiencing the proposed emotion (e.g., I am a calm person) or behavior (e.g., I am restless and fidgety) on a five-point Likert scale. Scores are summed for each item, with higher scores indicative of greater predisposition to arousal. The APS is a useful measure of individual differences in predisposition toward arousability and demonstrates adequate internal consistency.\(^{33}\)

**Statistical Analyses**

Data were analyzed using SPSS version 22.0 (IBM Corp., Armonk, NY). To assess the temporal stability of the FIRST, trait-anxiety (STAI-T), and arousability (APS), intraclass correlation coefficients (ICCs) [with 95% confidence interval (CI)] of the total scores of each scale were computed for two 6-mo intervals (baseline to 6 mo and 6 to 12 mo). Two 6-mo test-retest intervals were deemed appropriate to evaluate the robustness of the conclusions; however, ICCs for all assessments were also computed.
An ICC is a measure of agreement, which allows for fixed and random effects and provides an estimate of how reliably stable the FIRST scores are in participants with and without insomnia. ICCs were computed using a two-way random effect model and average measures were calculated [ICC (2, k)]. ICCs range from 0 to 1, with 0 indicative of high variability between baseline and follow-up estimates, and 1 indicative of low variability (i.e., high stability/reproducibility) between baseline and follow-up estimates. ICCs were classified as very poor stability (0-0.2), fair stability (0.3-0.4), moderate stability (0.5-0.6), strong stability (0.7-0.8), or excellent stability (0.9-1.0).

To evaluate the temporal stability of the FIRST with the stability of other trait-measuring scales, the ICC of the FIRST was compared with the 95% CI of trait-anxiety and arousability. Temporal stability between scales (FIRST and trait-anxiety, FIRST and arousability) was deemed comparable if the ICC of the FIRST fell within the 95% CI of trait-anxiety or arousability. If the ICC of the FIRST fell outside the 95% CI of trait-anxiety or arousability, there was a significant difference in temporal stability between scales (thus, not comparable). Statistical significance was set at 0.05 (two-tailed).

RESULTS

A majority of the participants were White (97.2%), married (57.3%), and worked full time (48.1%). Baseline FIRST measures ranged from 9 to 36 for all participants (insomnia syndrome: mean = 26.2, SD = 6.5; insomnia symptoms: mean = 23.5, SD = 6.6; good sleepers: mean = 19.3, SD = 5.7). Average baseline trait-anxiety for participants with insomnia syndrome, insomnia symptoms, and good sleepers were 45.2 (SD = 11.6), 38.7 (SD = 9.8), and 32.0 (SD = 8.2), respectively. Mean baseline arousability for participants with insomnia syndrome, insomnia symptoms, and good sleepers were 33.5 (SD = 8.2), 30.0 (SD = 7.0), and 27.5 (SD = 7.0), respectively.

Table 1 shows all ICCs for the FIRST, trait-anxiety, and arousability. Based on criteria by Shrout and Fleiss, the FIRST yielded strong temporal stability for baseline to
6-month (ICC = 0.81, p < 0.001) and for 6- to 12-month intervals (ICC = 0.74, p < 0.001) in individuals with insomnia syndrome. Similar results were observed among those with insomnia symptoms (ICCs range: 0.78-0.82, p’s < 0.001) and among good sleepers (ICCs range: 0.81-0.84, p’s < 0.001). ICCs were strong to excellent across all sleep groups at both 6-mo intervals for trait-anxiety (ICCs range: 0.86-0.94, p’s < 0.001) and for arousability (ICCs range: 0.82-0.91, p’s < 0.001).

ICCs across all assessments revealed strong stability for the FIRST in those with an insomnia syndrome (ICC = 0.84, 95% CI: 0.78, 0.88), insomnia symptoms (ICC = 0.85, 95% CI: 0.80, 0.89), and good sleep (ICC = 0.87, 95% CI: 0.86, 0.89, p’s < 0.001). Excellent stability was observed across all assessments for trait-anxiety among those with an insomnia syndrome (ICC = 0.94, 95% CI: 0.92, 0.96), insomnia symptoms (ICC = 0.91, 95% CI: 0.88, 0.93) and good sleep (ICC = 0.92, 95% CI: 0.91, 0.93, p’s < 0.001). Similar results were observed for arousability among those with an insomnia syndrome (ICC = 0.93, 95% CI: 0.91, 0.95), insomnia symptoms (ICC = 0.89, 95% CI: 0.86, 0.92) and good sleep (ICC = 0.89, 95% CI: 0.88, 0.91, p’s < 0.001).

Comparative analyses revealed that the temporal stability of the FIRST was significantly lower than trait-anxiety at both 6-mo intervals and across all sleep status groups. Temporal stability of the FIRST was somewhat comparable with arousability, but only among good sleepers for baseline to 6 mo and for participants with insomnia symptoms for the 6- to 12-mo interval. Temporal stability of the FIRST was significantly lower than that of arousability for the individuals with insomnia syndrome at both 6-mo intervals.

**DISCUSSION**

The aim of the current paper was to investigate the long-term stability of the FIRST over two 6-mo intervals and to compare its stability with the stability of other trait-measuring scales. Consistent with the hypothesis, the FIRST yielded strong temporal stability over
time. Coefficients across all participants were well within coefficients previously documented. The highest temporal stability was observed among good sleepers, followed by individuals with insomnia symptoms, and then syndrome. According to recognized criteria, the temporal stability of the FIRST was strong. However, when compared with more established scales, the FIRST was significantly lower, especially among participants with insomnia syndrome.

Unlike the trait-anxiety and arousability scales, which use more general statements (e.g., I am tense) to assess the trait of interest, the FIRST uses more specific scenarios (e.g., after having a bad day at work) to assess sleep reactivity. Still, temporal stability between the FIRST and the arousability scale were somewhat comparable in good sleepers and individuals with insomnia symptoms. This is consistent with data, which suggest participants with high FIRST scores are more aroused than those with low FIRST scores. It is possible that the FIRST may fluctuate as a function of the stressors perceived by an individual, and thus not reach the same level of stability as trait-anxiety. Further, those with chronic insomnia may be more habituated to their sleep disturbances, and may not necessarily attribute their sleep disturbances to specific stressful events as addressed by the FIRST. Conversely, those with good sleep may attribute their sleep disturbances to stressful events precisely because they are not accustomed to sleep disturbances.

The current findings provide support that the FIRST is a temporally reliable, stable scale over 6-mo intervals. That being said, a limitation of the current study was the possibility that participants became familiar with the scales over time; however, it could be argued that the intervals between assessments were relatively long enough for participants to answer without recall bias. Further, the results from the 6- to 12-mo interval corroborated the robustness of the results, suggesting substantial stability of the FIRST. Strengths of the current study include the large population-based sample of
participants with verified (at all assessments) chronic insomnia syndrome, insomnia symptoms, and good stable sleep. Sleep status groups were based on established criteria using validated criteria and questionnaires. Moreover, to the best of our knowledge, this is the first study to establish long-term reliability in the FIRST in a large sample.

Future research is needed to further corroborate the stability and trait-like measures of sleep reactivity with additional physiological, behavioral, and personality measures. Indeed, evidence suggests the FIRST is associated with rumination, pathological metacognitive beliefs about sleep, trait-anxiety, arousability, and neuroticism. Thus, it is worth investigating whether different vulnerable phenotypes are influenced or triggered by different psychological (e.g., personality), physiological (e.g., stress), or behavioral factors (e.g., coping styles).

Additionally, it would be valuable to investigate the temporal stability of the FIRST cut-offs proposed by Kalmbach et al. to further demonstrate its clinical utility and to better identify more efficient treatment and preventive campaigns (e.g., when screening should be offered). For example, reliably detecting individuals at risk for incident and persistent insomnia and providing early intervention to reliably evaluate long-term health outcomes. Relatedly, it would be important to investigate the long-term stability of the FIRST across the most common trajectories of insomnia (e.g., incidence, persistence, remission, and relapse), as this could be used to detect vulnerable individuals. In summary, the FIRST yields high temporal stability in individuals with and without insomnia.
ABBREVIATIONS

APS, Arousal Predisposition Scale

DSM, Diagnostic and Statistical Manual of Mental Disorders

CI, confidence interval

FIRST, Ford Insomnia Response to Stress Test

ICC, intraclass correlation coefficient

ICD, International Classification of Diseases

ISI, Insomnia Severity Index

PSQI, Pittsburgh Sleep Quality Index

STAI-T State-Trait Anxiety Inventory, Trait Anxiety Scale
Table 1—Intraclass correlation coefficients for the FIRST, STAI-T, and APS across two 6-mo intervals in participants with and without insomnia.

<table>
<thead>
<tr>
<th></th>
<th>Ford Insomnia Response to Stress Test (FIRST)</th>
<th>State-Trait Anxiety Inventory, Trait Anxiety Scale (STAI-T)</th>
<th>Arousal Predisposition Scale (APS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Insomnia Syndrome (n = 159)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline to 6 mo</td>
<td>0.81 (0.73, 0.85)</td>
<td>0.92 (0.88, 0.94)</td>
<td>0.91 (0.87, 0.93)</td>
</tr>
<tr>
<td>6 to 12 mo</td>
<td>0.74 (0.65, 0.81)</td>
<td>0.94 (0.92, 0.96)</td>
<td>0.91 (0.88, 0.93)</td>
</tr>
<tr>
<td><strong>Insomnia Symptoms (n = 152)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline to 6 mo</td>
<td>0.78 (0.70, 0.84)</td>
<td>0.86 (0.81, 0.90)</td>
<td>0.88 (0.83, 0.91)</td>
</tr>
<tr>
<td>6 to 12 mo</td>
<td>0.82 (0.75, 0.87)</td>
<td>0.89 (0.85, 0.92)</td>
<td>0.82 (0.75, 0.87)*</td>
</tr>
<tr>
<td><strong>Good sleeper (n = 811)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline to 6 mo</td>
<td>0.81 (0.78, 0.83)</td>
<td>0.87 (0.85, 0.89)</td>
<td>0.83 (0.80, 0.85)*</td>
</tr>
<tr>
<td>6 to 12 mo</td>
<td>0.84 (0.82, 0.86)</td>
<td>0.90 (0.89, 0.92)</td>
<td>0.87 (0.86, 0.89)</td>
</tr>
</tbody>
</table>

*Comparably temporal stability with the FIRST.*
REFERENCES


35. Lahey MA, Downey RG, Saal FE. Intraclass correlations: there’s more there than meets the eye. Psychol Bull 1983;93:586.
