SCIENTIFIC INVESTIGATIONS

Psychometric Evaluation of the Ford Insomnia Response to Stress Test (FIRST) in Early Pregnancy

Bizu Gelaye, PhD, MPH[1]; Qiu-Yue Zhong, MSc[1]; Yasmin V. Barrios, MPH[1]; Susan Redline, MD, MPH[1,2,3]; Christopher L. Drake, PhD[4]; Michelle A. Williams, ScD[1]

[1]Department of Epidemiology, Harvard School of Public Health, Boston, MA; [2]Department of Medicine, Brigham and Women's Hospital, Boston, MA;[3] Harvard Medical School, Boston, MA, [4]Sleep Disorders and Research Center, Henry Ford Hospital, Detroit, MI

Study Objectives: To evaluate the construct validity and factor structure of the Spanish-language version of the Ford Insomnia Response to Stress Test questionnaire (FIRST-S) when used in early pregnancy.

Methods: A cohort of 647 women were interviewed at ≤ 16 weeks of gestation to collect information regarding lifestyle, demographic, and sleep characteristics. The factorial structure of the FIRST-S was tested through exploratory and confirmatory factor analyses (EFA and CFA). Internal consistency and construct validity were also assessed by evaluating the association between the FIRST-S with symptoms of depression, anxiety, and sleep quality. Item response theory (IRT) analyses were conducted to complement classical test theory (CTT) analytic approaches.

Results: The mean score of the FIRST-S was 13.8 (range: 9–33). The results of the EFA showed that the FIRST-S contained a one-factor solution that accounted for 69.8% of the variance. The FIRST-S items showed good internal consistency (Cronbach α = 0.81). CFA results corroborated the one-factor structure finding from the EFA; and yielded measures indicating goodness of fit (comparative fit index of 0.902) and accuracy (root mean square error of approximation of 0.057). The FIRST-S had good construct validity as demonstrated by statistically significant associations of FIRST-S scores with sleep quality, antepartum depression and anxiety symptoms. Finally, results from IRT analyses suggested excellent item infit and outfit measures.

Conclusions: The FIRST-S was found to have good construct validity and internal consistency for assessing vulnerability to insomnia during early pregnancy.

Keywords: FIRST, stress-related sleep disturbance, pregnancy, Peru

Citation: Gelaye B, Zhong QY, Barrios YV, Redline S, Drake CL, Williams MA. Psychometric evaluation of the Ford Insomnia Response to Stress Test (FIRST) in early pregnancy. *J Clin Sleep Med* 2016;12(4):579–587.

INTRODUCTION

Pregnancy is associated with rapid physiological and biochemical changes that may alter sleep architecture, exacerbate sleep disordered breathing, and cause insomnia.[1] According to International Classification of Sleep Disorders (ICSD-3), insomnia is defined as a repeated difficulty with sleep initiation, duration, consolidation, or quality that occurs despite adequate opportunity and circumstances for sleep, and results in some form of daytime impairment.[2–4] Insomnia is highly prevalent during pregnancy, affecting 52% to 61% of women.[5] Insomnia during pregnancy has been associated with adverse outcomes including preeclampsia, gestational diabetes, and postpartum mood disorders.[6–9] The precise causes of insomnia are poorly understood, although available evidence suggests a combination of biological, psychological, and social determinants as causal factors.[2,3] Although insomnia can be the result of normal physiologic changes of pregnancy,[10] exposure to stressful life events or to less severe but more chronic daily struggles (e.g., occupational stress, marital discord) can serve as additional precipitating or exacerbating factors.[11,12] Individuals who are more prone to experience situational insomnia under stressful conditions may also be at greater risk for chronic insomnia.[13] Given (1) the high prevalence of insomnia in pregnancy; and (2) that insomnia is associated with substantial adverse

BRIEF SUMMARY

Current Knowledge/Study Rationale: The Ford Insomnia Response to Stress Test (FIRST) questionnaire has emerged as a broadly used instrument that assesses individual's vulnerability to situational insomnia under stressful conditions. Despite the increased use of the FIRST questionnaire to assess stress-related sleep reactivity in men and non-pregnant women, the instrument has not been used to assess stress-related sleep reactivity in pregnant women.

Study Impact: Our results indicate that the FIRST was found to have good construct validity and internal consistency for assessing vulnerability to insomnia during early pregnancy. Given (1) the high prevalence of insomnia in pregnancy; and (2) that insomnia is associated with substantial adverse maternal and newborn outcomes, the findings may have important clinical implications in maternal-fetal medicine.

maternal and newborn outcomes,[1] we reasoned that identifying women at high risk of developing insomnia may have important clinical implications in maternal-fetal medicine.

Developed by Drake et al., the Ford Insomnia Response to Stress Test (FIRST) questionnaire has emerged as a broadly used instrument that assesses individual's vulnerability to situational insomnia under stressful conditions.[14,15] Despite the increased use of the FIRST questionnaire to assess stress-related sleep reactivity in men and non-pregnant women,[14,16,17] the instrument has not, to our knowledge, been used to assess

Figure 1—Distribution of total score of the Spanish-language version of the Ford Insomnia Response to Stress Test (FIRST-S) among pregnant Peruvian women (n = 647).
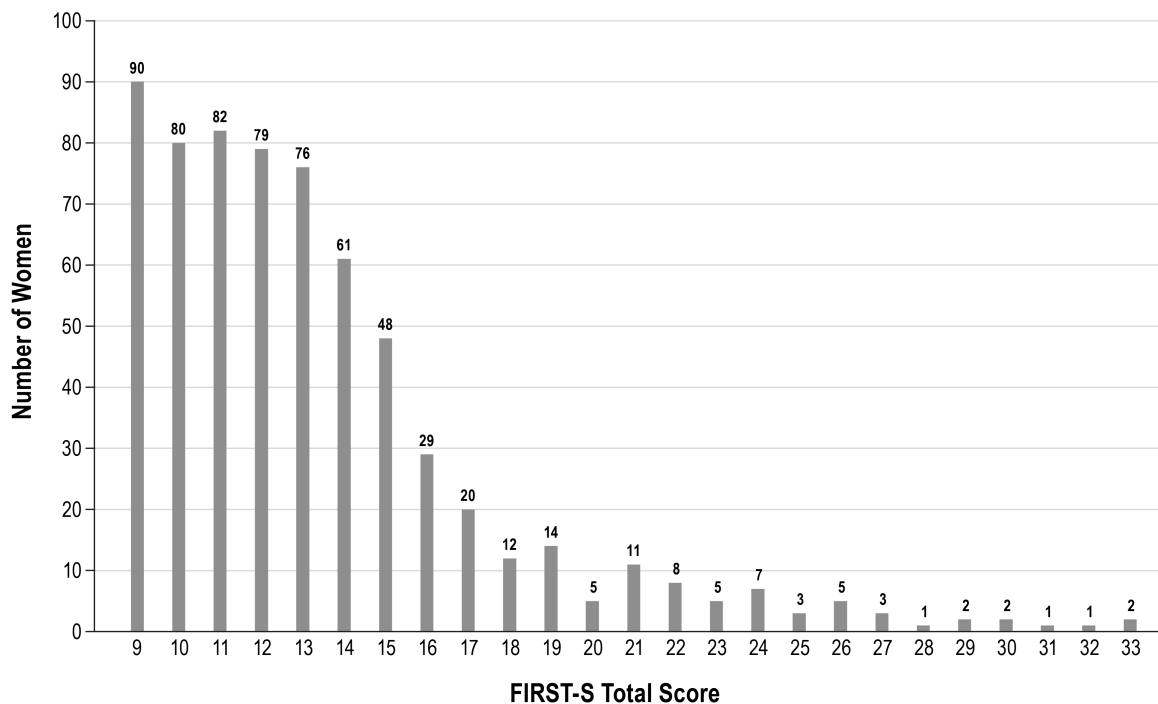
stress-related sleep reactivity in pregnant women. Of note, given that prior studies have shown that sleep, mood, and anxiety disorders are highly prevalent among low-income pregnant women in Peru,[18,19] we sought to evaluate the psychometric properties, construct validity and factor structure of the Spanish-language version of FIRST questionnaire (FIRST-S) in early pregnancy (**Figure 1**).

METHODS

Study Population

This analysis uses data initially gathered for the Pregnancy Outcomes, Maternal and Infant Study (PrOMIS) Cohort Study. The study population was drawn from pregnant women receiving prenatal care at the Instituto Nacional Materno Perinatal (INMP) in Lima, Peru, between February 2012 and March 2014. The INMP is the primary referral hospital for maternal and perinatal care and is operated by the Peruvian Ministry of Health. Pregnant women with a gestational age ≤ 16 weeks and who were between the ages of 18 and 49 years were eligible for inclusion. All participants provided written informed consent and all study procedures were approved by institutional review boards from the INMP, Lima, Peru, and the Human Research Administration Office at Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA.

Data Collection

During their initial prenatal care visit, enrolled subjects were asked to participate in in-person interviews conducted by trained research interviewers using a standardized structured questionnaire. The questionnaire was designed to collect information regarding maternal age, educational attainment, marital status, employment status during pregnancy, personal reproductive and medical histories, as well sleep habits during early pregnancy. The questionnaire, originally written in English, was translated into Spanish by a team of native Spanish speakers (including coauthor, YVB) with experience in sleep research. To ensure proper expression and conceptualization of terminologies in local contexts, the translated version was back-translated and modified until the back-translated version was comparable with the original English version.

Study Population

The study population for this report is derived from participants who enrolled in the PrOMIS Cohort Study between October 2013 and February 2014. Of the 724 participants approached, 662 participants completed the interview (91.4% response rate). A total of 15 participants excluded because of missing or incomplete information pertinent to this analysis. Hence, a total of 647 participants were included in the present analysis. In order to evaluate the construct validity and factor structure of FIRST-S when used in a general obstetric population (i.e., our research target population),[20] we did not exclude participants on the basis of their diagnoses of other sleep disorders.

Ford Insomnia Response to Stress Test (FIRST)

The FIRST is a standardized questionnaire that has been shown to be a sensitive measure of vulnerability to sleep disturbance and to have high reliability (test-retest reliability

coefficient = 0.92).[14] The FIRST includes 9 items asking about the likelihood of sleep disruption due to specific stressful situations and more broadly described periods of stress occurring during the day or evening. The possible responses and corresponding score included: not likely = 1, somewhat likely = 2, moderately likely = 3, and very likely = 4. The total score ranges from 9 to 36. High scores on the FIRST indicate greater vulnerability to sleep disruption.[21] Consistent with prior studies in the present study, we used the median score (12) to define high and low FIRST score groups.[16,21–24]

Patient Health Questionnaire-9 (PHQ-9)

The PHQ-9 was used to evaluate antepartum depression.[25] The PHQ-9 is a 9-item questionnaire with demonstrated reliability and validity for assessing depressive disorders among a diverse group of obstetrics-gynecology patients[26] and in Spanish-speaking women.[27,28] The PHQ-9 instrument asks respondents to rate the relevancy of each statement comprising emotional, cognitive, and functional somatic symptoms over the past 2 weeks on a 4-point scale (a) never; (b) several days; (c) more than half the days; or (d) nearly every day. The PHQ-9 total score is the sum of scores for the 9 items for each woman, and ranged from 0–27. Women were assigned to 1 of 5 depressive symptom categories based on total PHQ-9 score: (a) no depressive symptoms (0–4); (b) mild (5–9); (c) moderate (10–14); (d) moderately severe (15–19); and (e) severe (20–27) depressive symptoms. For the purpose of this study, we defined presence of antepartum depression based upon total PHQ-9 score, (a) no depressive symptoms (0–9) and (b) antepartum depression (10–27).[25] A meta-analysis of 14 studies supports the use of a PHQ-9 score of ≥ 10 to classify subjects with major depressive disorder.[29]

Pittsburgh Sleep Quality Index (PSQI)

The PSQI is a 19-item standardized self-reported questionnaire designed to assess sleep quality over the past month.[30] The PSQI has 7 sleep components: sleep duration, disturbance, latency, habitual sleep efficiency, use of sleep medicine, daytime dysfunction due to sleepiness, and overall quality of sleep. Each component produced a score ranging from 0 to 3, where a score of 3 indicates the highest level of dysfunction. A global sleep quality score is obtained by summing the individual component scores (range 0 to 21), with higher scores indicative of poorer sleep quality during the previous month. Participants with global scores > 5 were classified as poor sleepers.[30] Those with a score ≤ 5 were classified as good sleepers. This classification system is consistent with prior studies conducted globally including Peru.[31,32] We have demonstrated that the Spanish-language version of the PSQI instrument to have good construct validity among pregnant women in Peru.[33]

Generalized Anxiety Disorder 7 (GAD-7)

The GAD-7 is a 7-item questionnaire developed to identify probable generalized anxiety disorder (GAD) symptom severity.[34] The GAD-7 items include: (1) nervousness; (2) inability to stop worrying; (3) excessive worry; (4) restlessness; (5) difficulty in relaxing; (6) easy irritation; and (7) fear of something awful happening. Similar to the PHQ-9, the GAD-7 instrument asks respondents to rate the relevancy of each statement over the past 2 weeks on a 4-point scale: (a) never; (b) several days; (c) more than half the days; or (d) nearly every day, scored as 0, 1, 2, and 3, respectively. The GAD-7 total score (i.e., sum of scores for the 7 items) ranges from 0–21. In accordance with the original validation study, we used a cutoff score ≥ 10 to identify probable GAD.[34]

Other Covariates

Participants' age was categorized as follows: 18–20, 20–29, 30–34, and ≥ 35 years. Other sociodemographic variables were categorized as follows: maternal ethnicity (Mestizo vs. others); educational attainment (≤ 6, 7–12, and > 12 completed years of schooling); marital status (married and living with partner vs. others); employment status (employed vs. not employed); access to basic foods—a proxy for socioeconomic status (very hard/hard, somewhat hard, not very hard); parity (nulliparous vs. multiparous); planned pregnancy (yes vs. no); self-reported health in the last year (good vs. poor); and gestational age at interview.

Statistical Analysis

We first explored frequency distributions of demographic and lifestyle characteristics of the study participants. Participants' characteristics were summarized using means (± standard deviation [SD]) for continuous variables and counts and percentages for categorical variables with normal distributions. For variables with non-normal distributions as assessed using the Shapiro-Wilk test, we report median (interquartile range [IQR]). We assessed the internal consistency of the FIRST questionnaire using Cronbach α and item-total correlations. We report the Spearman correlation as a measure of the direction and strength of item-total correlations. Further, we explored the factor structure of the FIRST questionnaire using both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) approaches. Prior to performing EFA, we assessed the suitability of the data for performing factor analysis. This assessment showed that it was appropriate to proceed with factor analysis (Bartlett test of sphericity, p value < 0.001; and the Kaiser-Meyer-Olkin measure of sampling adequacy = 0.84). We conducted the EFA using principal component analysis with orthogonal rotation. We used the scree plot, presenting eigenvalues associated with each factor, to determine factor structure (**Figure 2**). Factors with eigenvalues > 1 were assumed to be meaningful and were retained for rotation. Rotated factor loadings of > 0.4 were considered sufficient, while items with factor loadings ≥ 0.4 on more than one factor were considered cross loading. To complement our EFA and to evaluate the model fit, we conducted CFA using maximum likelihood estimation approaches. In order to assess model fit, we calculated the (1) root mean square error of approximation (RMSEA) along with 95% confidence interval (95% CI), (2) comparative fit index, and (3) Tucker-Lewis index. Although there is an absence of consensus concerning the cutoff for goodness of fit, we elected to use the following recommended criteria for evidence of reasonably good fit: (1) SRMR values ≤ 0.08; (2) RMSEA values ≤ 0.06 or below; (3) comparative fit index ≥ 0.95.[35]

Figure 2—Scree plot of the Spanish-language version of the Ford Insomnia Response to Stress Test (FIRST-S) among pregnant Peruvian women (n = 647).
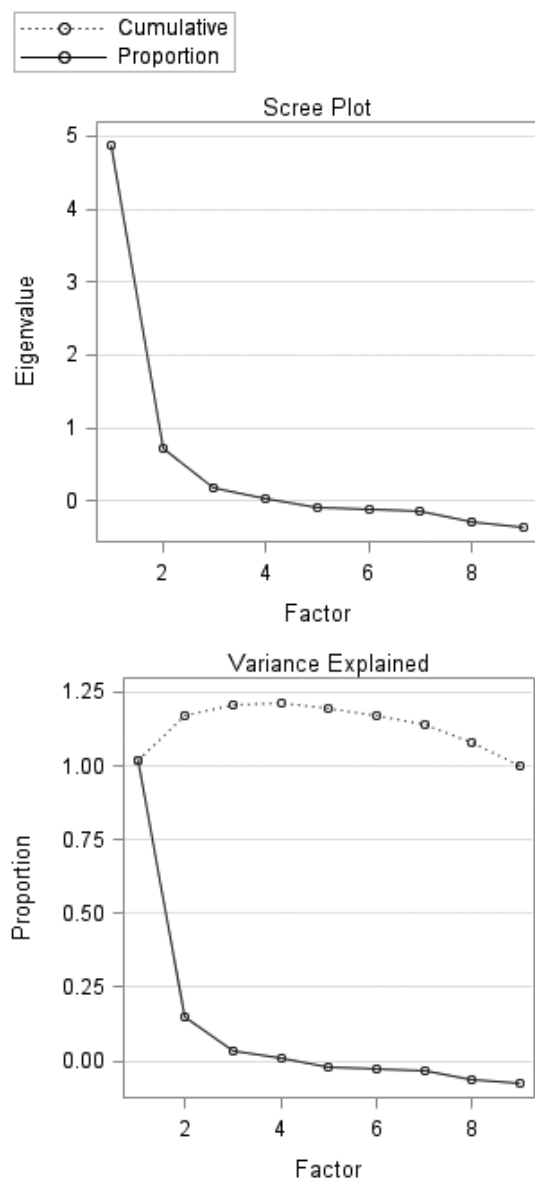
Poor Sleep Quality, Depressive Disorder, and Anxiety Disorder

We used Wilcoxon-Mann-Whitney test to compare mean PHQ-9, GAD-7, and PSQI scores for those with high FIRST score and those with low FIRST scores. Next, we used Spearman rank-order correlation to evaluate the correlation coefficients between the scores of the FIRST, PHQ-9, GAD-7, and PSQI controlling for maternal age to identify the association between stress-related sleep reactivity with depression, anxiety, and with severity of insomnia symptoms. We used unadjusted and multivariable-adjusted logistic regression models to calculate odds ratios (ORs) and 95% confidence intervals (CIs) of the association between sleep reactivity with poor sleep quality, antepartum depression, and generalized anxiety disorder. Additionally, we performed

a Rasch item response theory (IRT) analysis to evaluate the extent to which items from the Spanish version of the FIRST questionnaire are reliable and valid in detecting stress-related sleep reactivity in the present study population. The Rasch IRT is a one-parameter logistic item response model that provides a mathematical framework against which dichotomous data can be compared, and is used to determine the fitness and internal validity of screening instruments.[36] The Rasch IRT model assumes that the items within the questionnaire are unidimensional, measure the same construct and are independent of one another. The first two assumptions are assessed by mean square infit and outfit statistics, which measure the difference in the expected and the actual responses. The infit statistic is a weighted mean square residual value that is more sensitive to the unexpected response of an individual's ability level. The outfit statistic is the usual unweighted mean square residual and is more sensitive to unexpected observations or outliers. As a general rule mean square infit and outfit values between 0.7–1.3 indicate an acceptable fit to the model.[37] Fit statistics > 1.5 and < 0.5 indicate too much and too little variation in response patterns and should be considered for removal from the instrument to improve fit. In addition to the fit of the data and the model,[38] we also evaluated the person and item or Wright map to evaluate the hierarchy of the item difficulties. All statistical analyses were performed using SAS 9.3 (SAS Institute, Cary, NC, USA), Stata 11.0 (Statacorp, College Station, TX), Winsteps 3.80.0 (Chicago, IL), and R 3.1.0 using the "irt" package. The level of statistical significance was set at p values < 0.05, and all tests were two-tailed.

RESULTS

Sociodemographic and lifestyle characteristics of study participants are summarized in **Table 1**. A total of 647 participants (mean age = 28.7 years, SD = 6.5 years) participated in the study. The majority of participants were between the ages of 20–29 years (53.5%), and married/living with partner (79.9%). Approximately 49% of participants reported having difficulty accessing basic foods. Fifty percent of the cohort was employed during pregnancy. Distributions of the sociodemographic and lifestyle characteristics according to participants' high and low FIRST-S scores are also presented in **Table 1**. The mean score of the FIRST-S was 13.8 (SD 4.3; range: 9–33). A total of 316 participants was classified as having a high FIRST-S score. No significant difference in sociodemographic lifestyle characteristics was noted by stress-related sleep reactivity (according to high vs. low FIRST-S score).

The reliability coefficient, Cronbach α for the FIRST-S questionnaire total score was 0.81 (**Table 2**). The correlations between 9 items of the FIRST-S questionnaire and the total scores ranged from 0.33 to 0.62, and all correlations were statistically significant (all 2-tailed p values < 0.01). "Sleep disturbance after receiving bad news during the day" (r = 0.62) and "after an argument" (r = 0.62) were the 2 items with the highest item-total correlation coefficients. Conversely, "after watching a frightening movie or TV show" (r = 0.33) and "before having to speak in public" (r = 0.35) had the lowest item-total correlations.

Table 1—Characteristics of participants with high and low scores* on the Spanish-language version of the Ford Insomnia Response to Stress Test (FIRST-S) among pregnant Peruvian women (n = 647).

Selected Variables	All Participants (n = 647)	High FIRST-S Scores (FIRST-S > 12) (n = 316)	Low FIRST-S Scores (FIRST-S ≤ 12) (n = 331)	p value**
Maternal age, years, mean ± SD	28.7 ± 6.5	29.3 ± 6.8	28.2 ± 6.3	0.06
Maternal age, years, n (%)				
18–20	19 (2.9)	10 (3.2)	9 (2.7)	0.24
20–29	346 (53.5)	158 (50.0)	188 (56.8)	
30–34	144 (22.3)	71 (22.5)	73 (22.1)	
≥ 35	138 (21.3)	77 (24.4)	61 (18.4)	
Education, years, n (%)				
≤ 6	26 (4.0)	14 (4.4)	12 (3.6)	0.75
7–12	331 (51.2)	164 (51.9)	167 (50.5)	
> 12	289 (44.7)	137 (43.4)	152 (45.9)	
Mestizo ethnicity, n (%)	483 (74.7)	227 (71.8)	256 (77.3)	0.11
Married/living with partner, n (%)	517 (79.9)	252 (79.8)	265 (80.1)	0.99
Employed, n (%)	323 (49.9)	156 (49.4)	167 (50.5)	0.78
Access to basic foods, n (%)				
Hard	316 (48.8)	157 (49.7)	159 (48.0)	0.68
Not very hard	331 (51.2)	159 (50.3)	172 (52.0)	
Nulliparous, n (%)	302 (46.7)	160 (50.6)	142 (42.9)	0.48
Planned pregnancy, n (%)	266 (41.1)	127 (40.2)	139 (42.0)	0.62
Early pregnancy body mass index, kg/m^2, n (%)				
< 18.5	15 (2.3)	8 (2.5)	7 (2.1)	0.42
18.5–24.9	297 (45.9)	147 (46.5)	150 (45.3)	
25–29.9	238 (36.8)	121 (38.3)	117 (35.4)	
≥ 30	86 (13.3)	35 (11.1)	51 (15.4)	
Gestational age at interview, mean ± SD	9.2 ± 3.6	9.1 ± 3.7	9.2 ± 3.5	0.79

*Low and high FIRST-S scores were defined based on the median score (12) of the FIRST-S among this study population. **p value was calculated using Wilcoxon-Mann-Whitney test for continuous variables and χ^2 test for categorical variables.

Table 2—Item characteristics, item-total correlation, α if item deleted of the Spanish-language version of the Ford Insomnia Response to Stress Test (FIRST-S) among pregnant Peruvian women (n = 647).

Items	Mean	SD	Corrected Item-Total Correlation	α if Item Deleted
1. Before an important meeting the next day	1.42	0.68	0.39	0.80
2. After a stressful experience during the day	1.48	0.73	0.59	0.78
3. After a stressful experience in the evening	1.43	0.75	0.58	0.78
4. After getting bad news during the day	1.90	0.86	0.62	0.77
5. After watching a frightening movie or TV show	1.38	0.77	0.33	0.81
6. After having a bad day at work	1.41	0.72	0.52	0.78
7. After an argument	1.69	0.89	0.62	0.77
8. Before having to speak in public	1.39	0.73	0.49	0.79
9. Before going on vacation the next day	1.28	0.64	0.35	0.80
FIRST total score	13.38	4.26	N/A	0.81*

*Overall Cronbach α. SD, standard deviation.

The results from EFA showed that a rotated factor solution for the FIRST-S (**Table 3**) contained one factor with an eigenvalue > 1.0, which accounted for 69.8% of the variance. Item factor loadings ranged from 0.35 to 0.70 (**Table 3**). The item sleep disturbance after a stressful experience during the day (0.70) was most strongly related to the underlying construct. "After a stressful experience in the evening," "after getting bad news during the day," and "after an argument" were the next sets of items strongly related to the underlying construct (i.e., stress-related sleep disturbance). For all of these variables the correlations between FIRST-S items and the construct were 0.68. The items that loaded least strongly

Table 3—The factor loadings in exploratory factor analysis* of the Spanish-language version of the Ford Insomnia Response to Stress Test (FIRST-S) among pregnant Peruvian women (n = 647).

Items	Current Study (n = 647)	Drake et al., 2004 (n = 104)
1. Before an important meeting the next day	0.44	0.56
2. After a stressful experience during the day	0.70	0.40
3. After a stressful experience in the evening	0.68	0.73
4. After getting bad news during the day	0.68	0.73
5. After watching a frightening movie or TV show	0.35	0.48
6. After having a bad day at work	0.57	0.68
7. After an argument	0.68	0.56
8. Before having to speak in public	0.55	0.42
9. Before going on vacation the next day	0.37	0.51

*Exploratory factor analysis was conducted using the maximum likelihood method; Bartlett test of sphericity: p < 0.0001; Kaiser-Meyer-Olkin measure of sampling adequacy: 0.84.

Table 4—Models evaluated for the Spanish-language version of the Ford Insomnia Response to Stress Test (FIRST-S) and corresponding fit indices using confirmatory factor analysis* among pregnant Peruvian women (n = 647).

	χ^2	df	CFI	RMSEA	GFI	C_{min} / DF
Model 1: 9 items	83.2297	27	0.9030	0.0568	0.9999	3.08
Model 2: 8 items (exclude item 9)	58.8224	20	0.6608	0.0548	0.8910	2.94
Model 3: 8 items (exclude item 5)	59.5780	20	0.6872	0.0553	0.8924	2.98
Model 4: 7 items (exclude item 5 and item 9)	52.1400	14	0.6670	0.0649	0.8948	3.72

*Confirmatory factor analysis was conducted using the weighted least square method. df, degree of freedom; CFI, comparative fit index (close to 1 is better, minimally > 0.90); RMSEA, root mean square error of approximation (good fit < 0.05); GFI, goodness of fit index (close to 1 is better, minimally > 0.90); C_{min} / DF, χ^2 divided by degree of freedom (acceptable range [2, 3], lower is better).

on the construct included sleep disturbance "before going on vacation the next day" (0.37) and "after watching a frightening movie or TV show" (0.35). In addition to our EFA, we carried out a CFA to examine the construct validity and to test whether the factorial structure of FIRST-S is composed of one or more factors (**Table 4**). The comparative fit index and Tucker-Lewis index values, indicators of goodness of fit, for FIRST-S were higher than 0.90. The goodness of fit index was close to one (0.99). The RMSEA value was 0.0568 indicating a reasonable error of approximation. Collectively, these values suggest that a one-factor model is the best fit for FIRST-S.

The PSQI, PHQ-9, and GAD-7 mean scores were compared according to high and low FIRST-S scores. Compared to participants with low FIRST-S scores, those with high FIRST-S scores were more likely to have higher mean PSQI scores (5.3 [SD 2.8] versus 3.7 [SD 2.2], p < 0.001), PHQ-9-score (9.4 [SD 6.0] versus 5.4 [SD 3.8], p < 0.001), and GAD-7 scores (7.9 [SD 5.6] versus 4.2 [SD 3.6], p < 0.001). As shown in **Table 5**, we calculated the odds ratio for high FIRST-S score in relation to poor sleep quality, antepartum depression, and anxiety disorder symptoms. After adjusting for possible confounding by age, parity, early pregnancy body mass index (BMI), and difficulty paying for the very basics compared to participants with low FIRST-S scores, those with high FIRST-S scores had more than 3-fold increased odds of poor sleep quality (OR = 3.33; 95% CI: 2.24–4.95), more than 5-fold increased odds of antepartum depression (OR = 5.22; 95% CI: 3.39–8.02),

and more than 4-fold increased odds of anxiety disorder symptoms (OR = 4.61; 95% CI: 3.17–6.69).

Table 6 indicates the item fit summary statistics for FIRST-S questionnaire using the Rasch IRT analysis. None of the items misfit the model according to criteria set *a priori* with infit mean square values ranging from 0.68 to 1.61 and outfit mean square values ranging from 0.74 to 1.19. Finally, the separation index of FIRST-S, the ability to discriminate between participants with stress-related sleep reactivity and those without, was within acceptable range.

DISCUSSION

This is the first study to evaluate the psychometric properties and factor structure the FIRST-S questionnaire in pregnancy. The internal consistency was found to be excellent (Cronbach α = 0.81). The results of the EFA confirmed the unidimensionality of the FIRST-S with acceptable factor loadings and item discrimination. CFA results corroborated the EFA with good comparative fit index (0.902) and root mean square error value (0.057). Concurrent validity was supported by analyses that showed that women with higher FIRST-S scores had statistically significantly elevated odds of poor sleep quality, antepartum depression, and generalized anxiety disorder, respectively. Finally, the Rasch IRT methods showed excellent item infit and outfit measures, further supporting the validity of the FIRST-S in pregnant and Spanish speaking population.

Table 5—Associations between the Spanish-language version of the Ford Insomnia Response to Stress Test (FIRST-S) with sleep quality, depression, and generalized anxiety disorder (n = 647).

	High FIRST-S scores* (FIRST-S > 12) (n = 316)		Low FIRST-S scores* (FIRST-S ≤ 12) (n = 331)		Unadjusted OR (95% CI)	Adjusted OR** (95% CI)
	n	%	n	%		
Sleep quality						
Good sleep quality (PQSI ≤ 5)	171	54.11	252	76.13	1.00 (Reference)	1.00 (Reference)
Poor sleep quality (PQSI > 5)	114	36.08	50	15.11	3.36 (2.29, 4.94)	3.33 (2.24, 4.95)
Depression						
No depression (PHQ-9 < 10)	196	62.03	293	88.52	1.00 (Reference)	1.00 (Reference)
Depression (PHQ-9 ≥ 10)	118	37.34	37	11.18	4.77 (3.16, 7.19)	5.22 (3.39, 8.02)
Generalized anxiety disorder						
No anxiety GAD-7 < 7)	152	48.10	265	80.06	1.00 (Reference)	1.00 (Reference)
Yes anxiety (GAD-7 ≥ 7)	157	49.68	60	18.13	4.56 (3.19, 6.53)	4.61 (3.17, 6.69)

Due to missing data, percentages may not add up to 100%. *Low and high FIRST-S scores were defined based on the median score (12) of the FIRST-S among this study population. **Adjusted for maternal age (years), parity (nulliparous vs. multiparous), early pregnancy body mass index (< 18.5; 18.5–24.9; 25–29.9; ≥ 30), and difficulty paying for the very basics (hard vs. not very hard). PSQI, Pittsburgh Sleep Quality Index; PHQ-9, the Patient Health Questionnaire 9-item; GAD-7, Generalized Anxiety Disorder 7-item.

Table 6—Item hierarchy and fit statistics of the Spanish-language version of the Ford Insomnia Response to Stress Test (FIRST-S) and other measures under the Rasch Rating Scale Model (n = 647).

Items	Item Difficulty in Logits	Model SE	Infit*		Outfit*	
			MnSq	Zstd	MnSq	Zstd
9. Before going on vacation the next day	0.70	0.08	1.51	5.0	1.19	1.7
5. After watching a frightening movie or TV show	0.28	0.07	1.61	6.5	1.42	4.1
8. Before having to speak in public	0.24	0.07	1.17	2.1	1.01	0.2
6. After having a bad day at work	0.18	0.07	1.02	0.3	0.92	−0.9
1. Before an important meeting the next day	0.15	0.07	1.16	2.0	1.19	2.0
3. After a stressful experience in the evening	0.10	0.07	0.96	−0.5	0.78	−2.7
2. After a stressful experience during the day	−0.04	0.07	0.84	−2.2	0.74	−3.3
7. After an argument	−0.58	0.06	0.82	−2.8	0.82	−2.6
4. After getting bad news during the day	−1.02	0.05	0.68	−5.9	0.74	−4.2
Mean	0.00	0.07	1.09	0.5	0.98	−0.6
SD	0.48	0.01	0.30	3.7	0.23	2.7

*Infit MnSq value and outfit MnSq value should range from 0.6 to 1.4, and from 0.5 to 1.7, respectively. SE, standard error, SD, standard deviation, MnSq, mean square, Zstd, z-standardized.

To our knowledge, there are only two published reports evaluating the validity of the FIRST. Overall our study results are consistent with these two prior studies. In the original validation Drake et al., in their population based study conducted in Detroit, Michigan, USA, found a one-factor solution with Cronbach α of 0.83 for the instrument. Furthermore, the authors noted that individuals with high median FIRST scores (≥ 20) were more likely to have poor sleep efficiency (p = 0.001) and long sleep latency (p = 0.001), as measured by nocturnal polysomnography consistent with good construct and discriminant validity. In a subsequent validation study of healthy and insomnia patients in Japan, Nakajima et al. found the Japanese version of the FIRST questionnaire (FIRST-J) to also be unidimensional, with Cronbach α of 0.89 and 0.87, respectively. The authors noted that that the FIRST-J score was statistically significantly correlated with State-Trait Anxiety Inventory (STAI) scale both in healthy and insomnia patients

(p < 0.01). Despite differences in population characteristics, sample size, and study settings, on balance, the findings of our study and those of Drake et al. and Nakajima et al. consistently document good construct validity and high internal consistency reliability. Our results additionally indicate the potential utility of using the FIRST-S as a screening instrument for stress-related sleep disturbance among pregnant women in early pregnancy. By employing rigorous methodological approaches, including using classical and item response theory approaches, our study adds important new information to the sparse research literature concerning the reliability and construct validity of FIRST-S that may be used among pregnant women, particularly Spanish-language speakers.

The FIRST-S demonstrated good item-total correlations. "Sleep disturbance after getting bad news during the day" (r = 0.62) and "after an argument" (r = 0.62) were the two items with the largest item-total correlation coefficients, while the

smallest component-total correlation coefficients were noted for "sleep disturbance after watching a frightening movie or TV show" (r = 0.33) and "before having to speak in public" (r = 0.35). These are not surprising given that watching a frightening movie or TV show and speaking in public are not common among Peruvian women. In instances where questions may not be applicable due to cultural and pragmatic reasons, some investigators have modified the specific content included in sleep disturbance questionnaires to fit the study population context.[32] For instance, in their validation study of the Berlin questionnaire among Indian adults Sharma et al.[39] replaced the item "sleepiness while driving a car" with "sleepiness while waiting for an appointment with doctor, while watching television, or while waiting in the queue to make payment of telephone or electricity bill." Similarly, Storfer-Isser et al. in their study among adolescents in the Cleveland Children's Sleep and Health Study replaced the item "in a car while stopped for few minutes in traffic" with "doing homework or taking a test."[40] Similar contextual modifications might be considered in future studies designed to assess the psychometric properties of FIRST in cross- cultural settings.

Strengths of the current study include a relatively large sample size, unique study population, and rigorous analytic techniques. Nevertheless, our study is not without limitations. First, criterion validity utilizing diagnostic gold standard, and test-retest reliability using repeated measures were not assessed. However, the criterion validity of the FIRST has been previously shown to be appropriate when compared with more invasive measures.[14] Second, we did not collect information regarding the persistence of stress-related sleep disturbance over time. We also did not examine the extent to which stress-related sleep disturbance was associated with other maternal sleep disorders such as sleep disordered breathing, periodic limb movements during sleep, narcolepsy, circadian rhythm sleep disorders, and parasomnias. Longitudinal studies with a more exhaustive assessments of maternal sleep disorders are warranted to follow up women during pregnancy and postpartum and to estimate bi-directional associations of sleep and psychiatric disorders in this population. Third, given our study was conducted in early pregnancy (average gestational age of 9.1 weeks), future studies are warranted to understand the stability of the psychometric properties of the FIRST questionnaire during the course of pregnancy. Finally, our study was restricted to pregnant women primarily from a low socioeconomic background with high burden of social, medical, and mental health problems.[41–43] Hence, readers should exercise caution when generalizing results from the present study to other populations. However, the consistency of our findings from studies done in the US and Japan suggest that FIRST-S may be useful for assessing stress-related sleep disturbances among Spanish speaking pregnant women.

In summary, our study provided strong evidence for the construct validity, internal consistency, and strong item-total correlations of the FIRST-S questionnaire in assessing stress-related sleep disturbance. The findings of our study will have important clinical and public health implications. This is the first study to evaluate the psychometric properties of the FIRST questionnaire (translated in Spanish) among pregnant women. Having a psychometrically well-characterized instrument such as the FIRST-S enhances the opportunity for identifying pregnant women with underlying vulnerabilities to sleep disturbances in pregnancy. In addition, use of cross-culturally validated FIRST-S will provide important information needed for enhancing early detection, diagnosis, and treatment of insomnia in pregnant women.

ABBREVIATIONS

CFA, Confirmatory Factor Analysis
CI, Confidence Interval
CTT, Classical Test Theory
EFA, Exploratory Factor Analysis
FIRST, Ford Insomnia Response to Stress Test
FIRST-S, Spanish-language version of the Ford Insomnia Response to Stress Test
GAD-7, Generalized Anxiety Disorder 7
ICD, International Classification of Sleep Disorders
INMP, Instituto Nacional Materno Perinatal
IQR, Interquartile Range
IRT, Item Response Theory
PHQ-9, Patient Health Questionnaire-9
PrOMIS, Pregnancy Outcomes, Maternal and Infant Study
PSQI, Pittsburgh Sleep Quality Index
RMSEA, Root Mean Square Error of Approximation
STAI, State-Trait Anxiety Inventory

REFERENCES

1. Pien GW, Schwab RJ. Sleep disorders during pregnancy. Sleep 2004;27:1405–17.

2. Simon GE, VonKorff M. Prevalence, burden, and treatment of insomnia in primary care. Am J Psychiatry 1997;154:1417–23.

3. Institute of Medicine, Committee on Sleep Medicine and Research. In: Colten HR, Altevogt BM, eds. Sleep disorders and sleep deprivation: an unmet public health problem. Washington, DC: National Academy of Sciences Press, 2006.

4. American Academy of Sleep Medicine. International classification of sleep disorders, 3rd ed. Darien, IL: American Academy of Sleep Medicine, 2014.

5. Abbott SM, Attarian H, Zee PC. Sleep disorders in perinatal women. Best Pract Res Clin Obstet Gynaecol 2014;28:159–68.

6. Qiu C, Enquobahrie D, Frederick IO, Abetew D, Williams MA. Glucose intolerance and gestational diabetes risk in relation to sleep duration and snoring during pregnancy: a pilot study. BMC Women's Health 2010;10:17.

7. Edwards N, Blyton CM, Kesby GJ, Wilcox I, Sullivan CE. Pre-eclampsia is associated with marked alterations in sleep architecture. Sleep 2000;23:619–25.

8. Bei B, Milgrom J, Ericksen J, Trinder J. Subjective perception of sleep, but not its objective quality, is associated with immediate postpartum mood disturbances in healthy women. Sleep 2010;33:531–8.

9. Okun ML, Hanusa BH, Hall M, Wisner KL. Sleep complaints in late pregnancy and the recurrence of postpartum depression. Behav Sleep Med 2009;7:106–17.

10. Oyiengo D, Louis M, Hott B, Bourjeily G. Sleep disorders in pregnancy. Clin Chest Med 2014;35:571–87.

11. Pillai V, Roth T, Mullins HM, Drake CL. Moderators and mediators of the relationship between stress and insomnia: stressor chronicity, cognitive intrusion, and coping. Sleep 2014;37:1199–208.

12. Morin CM, Benca R. Chronic insomnia. Lancet 2012;379:1129–41.

13. Jarrin DC, Chen IY, Ivers H, Morin CM. The role of vulnerability in stress-related insomnia, social support and coping styles on incidence and persistence of insomnia. J Sleep Res 2014;23:681–8.

14. Drake C, Richardson G, Roehrs T, Scofield H, Roth T. Vulnerability to stress-related sleep disturbance and hyperarousal. Sleep 2004;27:285–91.

15. Drake CL, Jefferson C, Roehrs T, Roth T. Stress-related sleep disturbance and polysomnographic response to caffeine. Sleep Med 2006;7:567–72.

16. Nakajima S, Okajima I, Sasai T, et al. Validation of the Japanese version of the Ford Insomnia Response to Stress Test and the association of sleep reactivity with trait anxiety and insomnia. Sleep Med 2014;15:196–202.

17. Hall M, Thayer JF, Germain A, et al. Psychological stress is associated with heightened physiological arousal during NREM sleep in primary insomnia. Behav Sleep Med 2007;5:178–93.

18. Zhong Q, Gelaye B, Rondon M, et al. Comparative performance of Patient Health Questionnaire-9 and Edinburgh Postnatal Depression Scale for screening antepartum depression. J Affect Disord 2014;162:1–7.

19. Zhong Q, Gelaye B, Zaslavsky AM, et al. Diagnostic validity of the Generalized Anxiety Disorder - 7 (GAD-7) among pregnant women. PloS One 2015;10:e0125096.

20. Zhou XH, Obuchowski NA, Mcclish DK. Statistical methods in diagnostic medicine. Hoboken, NJ: John Wiley & Sons, 2011.

21. Drake CL, Scofield H, Roth T. Vulnerability to insomnia: the role of familial aggregation. Sleep Med 2008;9:297–302.

22. Drake CL, Friedman NP, Wright KP Jr, Roth T. Sleep reactivity and insomnia: genetic and environmental influences. Sleep 2011;34:1179–88.

23. Drake CL, Jefferson C, Roehrs T, Roth T. Stress-related sleep disturbance and polysomnographic response to caffeine. Sleep Med 2006;7:567–72.

24. Drake C, Richardson G, Roehrs T, Scofield H, Roth T. Vulnerability to stress-related sleep disturbance and hyperarousal. Sleep 2004;27:285–92.

25. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16:606–13.

26. Spitzer RL, Williams JB, Kroenke K, Hornyak R, McMurray J. Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. Am J Obstet Gynecol 2000;183:759–69.

27. Wulsin L, Somoza E, Heck J. The feasibility of using the Spanish PHQ-9 to screen for depression in primary care in Honduras. Prim Care Companion J Clin Psychiatry 2002;4:191.

28. Zhong Q, Gelaye B, Fann JR, Sanchez SE, Williams MA. Cross-cultural validity of the Spanish version of PHQ-9 among pregnant Peruvian women: a Rasch item response theory analysis. J Affect Disord 2014;158:148–53.

29. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med 2007;22:1596–602.

30. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28:193–213.

31. Sanchez S, Martinez C, Oriol R, et al. Sleep quality, sleep patterns and consumption of energy drinks and other caffeinated beverages among Peruvian college students. Health 2013;5:26–35.

32. Gelaye B, Lohsoonthorn V, Lertmeharit S, et al. Construct validity and factor structure of the pittsburgh sleep quality index and epworth sleepiness scale in a multi-national study of African, South East Asian and South American college students. PloS One 2014;9:e116383.

33. Zhong QY, Gelaye B, Sanchez SE, Williams MA. Psychometric properties of the Pittsburgh Sleep Quality Index (PSQI) in a cohort of Peruvian pregnant women. J Clin Sleep Med 2015;11:869–77.

34. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166:1092–7.

35. Hsieh CC, Maisonneuve P, Boyle P, Macfarlane GJ, Roberston C. Analysis of quantitative data by quantiles in epidemiologic studies: classification according to cases, noncases, or all subjects? Epidemiology 1991;2:137–40.

36. Brown TA. Confirmatory factor analysis for applied research. New York, NY: Guilford Press, 2006.

37. Bond T, Fox C. Applying the Rasch model: fundamental measurement in the human sciences. United Kingdom: Psychology Press, 2013.

38. Linacre JM. Winsteps Ministep Rasch-model computer programs. 2007.

39. Sharma SK, Vasudev C, Sinha S, Banga A, Pandey RM, Handa KK. Validation of the modified Berlin questionnaire to identify patients at risk for the obstructive sleep apnoea syndrome. Indian J Med Res 2006;124:281–90.

40. Storfer-Isser A, Lebourgeois MK, Harsh J, Tompsett CJ, Redline S. Psychometric properties of the Adolescent Sleep Hygiene Scale. J Sleep Res 2013;22:707–16.

41. Gelaye B, Lam N, Cripe SM, Sanchez SE, Williams MA. Correlates of violent response among Peruvian women abused by an intimate partner. J Interpers Violence 2010;25:136–51.

42. Miranda JJ, Lopez-Rivera LA, Quistberg DA, et al. Epidemiology of road traffic incidents in Peru 1973-2008: incidence, mortality, and fatality. PLoS One 2014;9:e99662.

43. Perales MT, Cripe SM, Lam N, Sanchez SE, Sanchez E, Williams MA. Prevalence, types, and pattern of intimate partner violence among pregnant women in Lima, Peru. Violence Against Women 2009;15:224–50.

SUBMISSION & CORRESPONDENCE INFORMATION

DISCLOSURE STATEMENT