

■ Original Article

Testing the Reliability and Validity of the Korean Version of the Pittsburgh Sleep Quality Index Using Fitbit Devices: A Cross-Sectional Analysis

Si-Yeon Lee¹, Ja-Eun Choi², Ji-Won Lee¹, Yaeji Lee^{3,4}, Jae-Min Park^{5,6,*}, Kyung-Won Hong^{2,*}

¹Department of Family Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

²R&D Division, Theragen Health Co. Ltd., Seongnam, Korea

³Department of Biostatistics and Computing, Yonsei University, Seoul, Korea

⁴Department of Family Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

⁵Department of Family Medicine, Uijeongbu Eulji Medical Center, Eulji University, Uijeongbu, Korea

⁶Department of Medicine, Graduate School of Medicine, Yonsei University, Seoul, Korea

Background: Sleep disorders and insomnia are prevalent worldwide, with negative health outcomes. The Pittsburgh Sleep Quality Index (PSQI) is a widely used self-report assessment tool for evaluating sleep quality, comprising seven subdomains. The Korean version of the PSQI (PSQI-K) has been tested for reliability and validity in small sample sizes but lacks large-scale validation using objective measures.

Methods: This study was conducted with 268 Korean adults attending health check programs. Participants completed the PSQI-K questionnaire and wore Fitbit devices (Fitbit Inc., USA) to ascertain sleep parameters. Reliability was analyzed using the Cronbach's α coefficient, and construct validity was determined through factor analysis. Criteria validity was assessed by correlating their index scores with Fitbit sleep parameters. We identified the optimal cutoff for detecting sleep disorders.

Results: The Cronbach's α coefficient was 0.61, indicating adequate internal consistency. Factor analysis revealed three factors, explaining 48.2% of sleep quality variance. The index scores were negatively correlated with Fitbit sleep efficiency, total sleep time, and number of awakenings ($P < 0.05$). The optimal cutoff point for identifying sleep disorder groups was ≥ 6 .

Conclusion: The PSQI-K demonstrated good reliability and validity when correlated with Fitbit sleep parameters, offering a practical screening tool for identifying sleep disorders among Korean adults. Cutoff scores can help identify patients for sleep interventions. However, further large-scale studies are required to validate these findings.

Keywords: Sleep; Sleep Quality; Reliability; Validity; Insomnia

Received: August 30, 2023, Revised: December 10, 2023, Accepted: December 12, 2023

*Corresponding Author: Jae-Min Park <https://orcid.org/0000-0001-8873-8832>

Tel: +82-31-951-2339, Fax: +82-31-951-3300, E-mail: milkcandy@eulji.ac.kr

*Corresponding Author: Kyung-Won Hong <https://orcid.org/0000-0002-3137-4609>

Tel: +82-31-288-1288, Fax: +82-31-288-1294, E-mail: kyungwon.hong@theragenhealth.com

*These authors contributed equally to this work as corresponding authors.

INTRODUCTION

The prevalence of sleep disorders and insomnia has surged across diverse age groups.¹⁾ The global sleep dynamics research estimates a 10%–20% occurrence of insomnia disorder within populations, with chronic symptoms in nearly half of the affected individuals.²⁾ Similarly, over one-fifth of adults experience sleep disorders in Korea.³⁾ Moreover, the incidence of insomnia has doubled over the last decade, emphasizing the need for effective assessment tools and interventions.⁴⁾

Accumulating evidence has underscored a correlation between compromised sleep quality and adverse health outcomes. Cardiovascular mortality, coronary heart disease, stroke, dementia, mental disorders, and metabolic syndrome are among the potential consequences of poor sleep quality.¹⁾ A recent study even highlights associations between sleep disturbances in patients with Parkinson's disease and alterations in white matter and functional networks, and cognitive impairments.⁵⁾

Evaluating sleep quality involves an interplay of factors, including sleep onset, maintenance, duration, and sense of revitalization upon awakening.⁶⁾ Sleep quality can be objectively and subjectively measured. Polysomnography, an objective method, is known to accurately collect sleep parameter data.⁷⁾ However, it is resource-intensive and has limited accessibility for clinicians, necessitating alternative approaches. Conversely, subjective self-rating questionnaires have the advantages of patient compliance, ease of administration, and reduced demands on healthcare personnel.

The Pittsburgh Sleep Quality Index (PSQI) is a widely utilized subjective sleep quality measure, encompassing seven key facets, including subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medication, and daytime dysfunction.⁸⁾ The Korean version of the PSQI (PSQI-K) used in Korea⁹⁾ and also in this study assigns scores of 0–3 to each facet, resulting in a total score range of 0–21. Higher scores indicate poorer sleep quality. Notwithstanding its frequent application, discrepancies in interpreting PSQI scores—such as the use of subdomain cutoffs—have yielded suboptimal assessments of sleep disorders.

To address this gap, a previous study examined the reliability and validity of the PSQI-K in a small sample of young adults based on data

from Fitbit devices (Fitbit Inc., San Francisco, CA, USA).¹⁰⁾ However, the study lacks robustness as only a limited number of young adults in their 20s were tested. This study therefore aimed to comprehensively examine the reliability and validity of the PSQI-K through a relatively large-scale analysis of Korean adults. This evaluation was enhanced by objective sleep assessments using Fitbit devices, thus providing a broader perspective on sleep quality.

METHODS

1. Study Overview and Participants

This cross-sectional study included participants attending medical health check programs at two institutions: the Department of Family Medicine of the Gangnam Severance Hospital and Severance Health Checkup, Seoul, Korea, between December 2019 and December 2021. Each participant completed the PSQI-K questionnaire.^{8,9)} The PSQI-K questionnaire is presented in Supplement 1. Each participant wore a Fitbit device on their wrist—one of the most common wireless physical activity trackers estimating sleep parameters and stages in the consumer market.^{11–13)} Among the 280 participants who voluntarily participated in this study, we excluded 12 individuals who had missing data or did not complete the questionnaire. Subsequently, 268 participants were included in the analysis. Written informed consent was obtained from each participant.

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki and approved by the Institutional Review Board of the Theragen Bio Institute (approval number: 700062-20190819-GP-006-02).

2. Measurement of Sleep Quality

The PSQI, developed by Buysse et al.,⁸⁾ was translated into Korean by Shon et al.⁹⁾ to measure sleep quality. It is a self-administered questionnaire consisting of 19 questions designed to assess sleep quality during the previous month. Each item in the PSQI-K is scored from 0 to 3, where 0 and 3 indicate positive and negative extremes on the Likert scale, respectively. The global PSQI-K score is the sum of these items, with scores ranging from 0–21.

Table 1. Item-total correlation, Cronbach's α , and factor structure of the Korean PSQI (N=268)

Item	Item mean \pm SD	Item-total correlation	Cronbach α if an item is deleted
PSQI1: subjective sleep quality	1.16 \pm 0.59	0.64	0.52
PSQI2: sleep latency	1.12 \pm 0.97	0.61	0.58
PSQI3: sleep duration	1.15 \pm 0.93	0.58	0.57
PSQI4: habitual sleep efficiency	0.59 \pm 0.89	0.64	0.54
PSQI5: sleep disturbance	1.05 \pm 0.52	0.54	0.55
PSQI6: use of sleep medication	0.09 \pm 0.45	0.24	0.62
PSQI7: daytime dysfunction	0.63 \pm 0.76	0.45	0.62
PSQI total	0.83 \pm 0.41		
Cronbach α	0.61		

PSQI, Pittsburgh Sleep Quality Index; SD, standard deviation.

3. Measurement of Sleep Pattern

Sleep efficiency, total sleep time, sleep latency, and awakenings index were measured using Fitbit—a reasonable and reliable tool used to estimate sleep parameters.^{12,13} Total sleep time implies the total time spent sleeping. Sleep latency refers to the time difference between lying in bed and actually falling asleep, and sleep efficiency is calculated as a percentage by dividing total sleep time by total bedtime. Number of awakenings during sleep refers to the number of awakenings from the start of sleep until immediately before the last awakening.¹⁴

4. Statistical Analysis

The internal consistency reliability of the PSQI-K was analyzed using Cronbach's α . For validity, a factor analysis was conducted using the varimax factor extraction and rotation method, and the number of factors was calculated using eigenvalues and factor loadings. Correlations between the total PSQI-K score and sleep efficiency, total sleep time, sleep latency, and awakenings index were analyzed using Spearman's rho correlation coefficient. The optimal cutoff, reflecting the sensitivity and specificity for identifying the "sleep disorder group" based on the measured PSQI-K, was analyzed with the receiver operating characteristic (ROC) curve based on the Fitbit's sleep efficiency index of $\leq 95\%$.

RESULTS

1. Characteristics of the Study Participants

The mean age of the participants was 46.5 years. Of the 268 participants, 62 (23.1%) were men and 206 (76.9%) were women.

2. Reliability

Table 1 presents the results of the reliability analysis of the PSQI-K. The criteria for calculating item-total correlation coefficients were set at ≥ 0.30 , and the internal consistency coefficient was set at ≥ 0.60 . Reliability testing of the PSQI-K revealed that the item-total correlation coefficient was 0.24–0.64, with all items except for the "use of sleep medication" item meeting the criterion of ≥ 0.30 . The Cronbach's α for the PSQI-K was 0.61.

3. Construct Validity

The results of the varimax principal component analysis and the factor rotation of the seven components of the PSQI-K tool are presented in

Table 2. Three factors were derived: the first included two items—sleep duration and habitual sleep efficiency—accounting for 20.2% of the variance. Factor 2 included three items—sleep latency, sleep disturbance, and use of sleep medication—accounting for 15.3% of the variance. Factor 3 included two items—subjective sleep quality and daytime dysfunction—accounting for 12.7% of the variance. The total explanatory power (variance) of the model was 48.2%. Bartlett's test of sphericity was 311.86 (degrees of Freedom=21, $P < 0.001$), indicating that the factor analysis model was appropriate. Additionally, the Kaiser-Meyer-Olkin measure of sampling adequacy was 0.60, confirming that the extracted factors adequately explained the variables.¹⁵

4. Criteria Validity

Table 3 presents the results of the correlation between the PSQI-K score and sleep efficiency, total sleep time, sleep latency, and awakenings index on Fitbit. The PSQI-K score was negatively correlated with Fitbit's sleep efficiency, total sleep time, and awakenings index ($r = -0.25$, $P < 0.001$; $r = -0.14$, $P = 0.024$; and $r = -0.16$, $P = 0.012$, respectively). However, PSQI-K score was not significantly associated with sleep latency ($P = 0.241$).

5. Sensitivity, Specificity, and the Cutoff

The results of the ROC curve analysis are presented in Table 4 and Fig-

Table 2. Factor analysis of the PSQI-K

Variable	Factor loading		
	1	2	3
Factor 1			
PSQI3: sleep duration	0.98		
PSQI4: habitual sleep efficiency	0.60		
Factor 2			
PSQI2: sleep latency		0.67	
PSQI5: sleep disturbance		0.54	
PSQI6: use of sleep medication		0.22	
Factor 3			
PSQI1: subjective sleep quality			0.47
PSQI7: daytime dysfunction			0.74
Eigenvalue	1.41	1.07	0.89
Percentage of variance explained	20.20	15.30	12.70
Total variance explained (%)	20.20	35.50	48.20

PSQI-K, Korean version of the Pittsburgh Sleep Quality Index. Kaiser-Meyer-Olkin=0.60. Bartlett's chi-square test=311.86 ($P < 0.001$).

Table 3. Correlation between the PSQI-K score and sleep efficiency, total sleep time, sleep latency, and awakenings index on the Fitbit

	Spearman's rho (P-value)				
	PSQI-K score	Sleep efficiency	Total sleep time	Sleep latency	Awakenings index
PSQI-K score	1.00				
Sleep efficiency	-0.25 (<0.001)	1.00			
Total sleep time	-0.14 (0.024)	0.10 (0.094)	1.00		
Sleep latency	-0.07 (0.241)	0.31 (<0.001)	0.18 (0.004)	1.00	
Awakenings index	-0.16 (0.012)	-0.15 (0.017)	0.49 (<0.001)	-0.05 (0.403)	1.00

PSQI-K, Korean version of the Pittsburgh Sleep Quality Index.

Table 4. PSQI-K performance

PSQI-K cutoff*	Sensitivity	Specificity
≥1	0.985	0.016
≥2	0.970	0.065
≥3	0.930	0.161
≥4	0.821	0.371
≥5	0.647	0.548
≥6	0.522	0.710
≥7	0.408	0.823
≥8	0.318	0.887
≥9	0.224	0.935
≥10	0.159	0.935
≥11	0.104	0.952
≥12	0.040	0.984
≥13	0.020	0.984
≥14	0.000	1.000
≥15	0.000	1.000

PSQI-K, Korean version of the Pittsburgh Sleep Quality Index.

*Scores higher than the cutoff indicate insomnia.

ure 1. The PSQI-K cutoff for determining sleep disorders was ≥ 6 —the most suitable for the sleep disorder group based on the sleep efficiency index of the Fitbit with a sensitivity of 0.52 and specificity of 0.71, and the area under the ROC curve was 0.66 (95% confidence interval, 0.58–0.73).

DISCUSSION

We verified the reliability and validity of the translated PSQI-K to evaluate subjective sleep quality and ascertain the optimal cutoff score for diagnosing sleep disorders in Korean adults. In this study, the reliability coefficient of the PSQI-K was 0.61, within the 0.6–0.8 standard for an adequate reliability coefficient. Cronbach's α is a measurement that calculates the internal consistency of an assessment instrument.¹⁶⁾ In this study, the reliability coefficient of the PSQI-K was 0.61, which was lower than that of 0.84 reported by Shon et al.⁹⁾ However, a Cronbach's alpha value between 0.6 and 0.8 is deemed acceptable.¹⁶⁻¹⁸⁾ Thus, this study confirmed the acceptable reliability of the tool. The difference in Cronbach's α between this study and that of Shon et al.⁹⁾ may be due to the use of polysomnography in their study,⁹⁾ whereas Fitbit was used in this study.

An exploratory factor analysis was conducted to verify the construct validity of the PSQI-K. Among the seven subfactors, those with an eigenvalue of ≥ 1.0 and factor loadings of ≥ 0.60 were categorized into three factors, explaining 48.2% of sleep quality. This finding differs from those of Shon et al.⁹⁾ where a two-factor division was obtained. Construct validity pertains to how effectively a set of indicators represents or reflects a concept that is not directly measurable. Our approach involved examining the covariations of factors and their loadings on latent variables through an exploratory factor analysis. The resultant structural model revealed a three-factor structure. By employing factor rotation, the initial matrix was transformed into a simple structure matrix, enhancing interpretability. To this process, the vari-

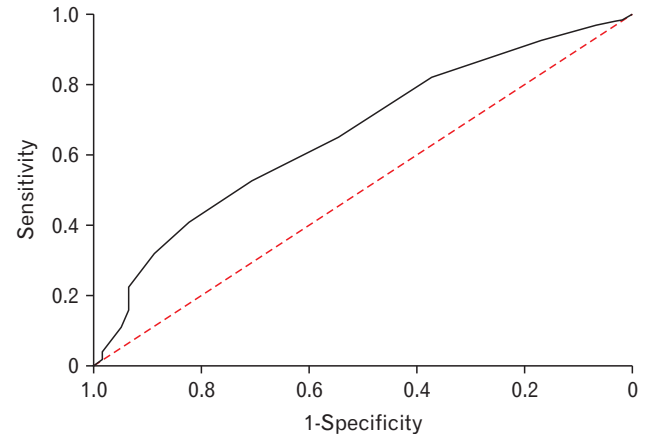


Figure 1. Receiver operating characteristic (ROC) curve for using the cutoff of ≥ 6 for the Korean version of the Pittsburgh Sleep Quality Index in insomnia as defined by sleep efficiency measured using the Fitbit device. The area under the ROC curve was 0.66 (95% confidence interval, 0.58–0.73).

max method was applied, assuming independence among factors and, consequently, zero inter-factor correlations. This method has the advantage of clearly distinguishing between factors, rendering interpretation more straightforward. In the evaluation of construct validity, the factors exhibited a distinct and well-defined latent meaning, facilitating a more straightforward interpretation.

The criterion validity of the PSQI-K was verified using indicators of sleep efficiency, total sleep time, sleep latency, and number of midnight awakenings measured using Fitbit trackers. The PSQI-K total score was weakly negatively correlated with Fitbit's sleep efficiency ($r=-0.25$), total sleep time ($r=-0.024$), and number of midnight awakenings ($r=-0.016$), indicating a lack of validity compared to the study by Shin and Kim¹⁰⁾ This may be due to the differences in the characteristics of the study participants. While the study by Shin and Kim¹⁰⁾ had a small sample size comprising only young adults working in hospitals, the present study had a larger and more diverse study population encompassing various ages.

Fitbit uses accelerometers and heart rate sensors to estimate sleep stages (light, deep, and rapid eye movement sleep). While these can provide a good approximation, they may not be as accurate as data collected through polysomnography, which involves a more extensive array of sensors, including electroencephalography.¹⁹⁾ Moreover, there is a potential for error in the results due to environmental stimuli or user factors; hence, the user regularity with using the Fitbit during sleep should also be considered.¹²⁾

However, recent advances in wearable devices and big data analytics have made it possible to collect and analyze lifestyle and environmental data. An integrated analysis of these data can predict individual health risks, based on which appropriate management measures through active participation in healthcare behaviors can be encouraged. These measures are expected to prevent or delay the onset of diseases, thereby minimizing economic losses for individuals and society.²⁰⁾ Therefore, this study, which evaluated the validity and reliabil-

ity of the PSQI-K—a diagnostic tool for sleep disorders—using a simple and user-friendly wearable device compared to polysomnography—an existing diagnostic tool—is expected to have positive significance.

The optimal cutoff for distinguishing sleep disorder groups based on the PSQI-K total score was 6 points, which is consistent with previous studies. Thus, Korean adults with a PSQI-K total score of ≥ 6 can be screened as “poor sleepers,” which may be a useful criterion for identifying sleep disorder groups in research and clinical practice.

This study has several limitations. We did not consider the history of sleep disorders or sleep medication types—two factors that should be accounted for in future studies. In addition, participants were selected from only two hospitals; therefore, the study population may not be representative of the general Korean population. Further reliability and validity should therefore be confirmed through studies involving representative populations. Despite this study’s limitations, we verified the reliability and validity of the PSQI-K using Fitbit for a relatively large number of participants of various ages, as compared to a previous study which included a small sample size comprising only young adults working in hospitals.¹⁰⁾

In conclusion, the PSQI-K, which is frequently used as a screening tool for sleep disorders in Korean adults, was highly correlated with the Fitbit tracker, a simple objective measurement tool, in terms of sleep efficiency, total sleep time, sleep latency, and number of awakenings. The study also verified a cutoff for the PSQI-K to identify “poor sleepers,” which can be used as a more convenient tool for screening patients for sleep interventions.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

FUNDING

This research was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry (IPET) through High Value-added Food Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (grant number: 321030051HD030).

ORCID

Si-Yeon Lee: <https://orcid.org/0009-0009-3858-2185>

Ja-Eun Choi: <https://orcid.org/0000-0002-4949-6348>

Ji-Won Lee: <https://orcid.org/0000-0002-2666-4249>

Yaeji Lee: <https://orcid.org/0000-0002-1411-1938>

Jae-Min Park: <https://orcid.org/0000-0001-8873-8832>

Kyung-Won Hong: <https://orcid.org/0000-0002-3137-4609>

SUPPLEMENTARY MATERIALS

Supplementary materials can be found via <https://doi.org/10.4082/kjfm.23.0166>.

kjfm.23.0166. Supplement 1. Korean version of the Pittsburgh Sleep Quality Index.

REFERENCES

- Sejbuk M, Mironczuk-Chodakowska I, Witkowska AM. Sleep quality: a narrative review on nutrition, stimulants, and physical activity as important factors. *Nutrients* 2022;14:1912.
- Buysse DJ. Insomnia. *JAMA* 2013;309:706-16.
- Korean Statistical Information Service. Survey on working condition: sleep disorder [Internet]. Daejeon: Statistics Korea; 2017-2020 [cited 2024 Mar 2]. Available from: https://kosis.kr/statHtml/statHtml.do?orgId=380&tblId=DT_380002_K006_6TH&conn_path=I2
- Chung S, Cho SW, Jo MW, Youn S, Lee J, Sim CS. The prevalence and incidence of insomnia in Korea during 2005 to 2013. *Psychiatry Investig* 2020;17:533-40.
- Chung SJ, Choi YH, Kwon H, Park YH, Yun HJ, Yoo HS, et al. Sleep disturbance may alter white matter and resting state functional connectivities in Parkinson’s disease. *Sleep* 2017;40:zsx009.
- Fabbri M, Beracci A, Martoni M, Meneo D, Tonetti L, Natale V. Measuring subjective sleep quality: a review. *Int J Environ Res Public Health* 2021;18:1082.
- Rundo JV, Downey R 3rd. Polysomnography. *Handb Clin Neurol* 2019;160:381-92.
- 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.
- Sohn SI, Kim DH, Lee MY, Cho YW. The reliability and validity of the Korean version of the Pittsburgh Sleep Quality Index. *Sleep Breath* 2012;16:803-12.
- Shin S, Kim SH. The reliability and validity testing of Korean version of the Pittsburgh Sleep Quality Index. *J Converg Inf Technol* 2020;10:148-55.
- Diaz KM, Krupka DJ, Chang MJ, Peacock J, Ma Y, Goldsmith J, et al. Fitbit: an accurate and reliable device for wireless physical activity tracking. *Int J Cardiol* 2015;185:138-40.
- de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2(TM) compared with polysomnography in adults. *Chronobiol Int* 2018;35:465-76.
- Burton E, Hill KD, Lautenschlager NT, Thøgersen-Ntoumani C, Lewin G, Boyle E, et al. Reliability and validity of two fitness tracker devices in the laboratory and home environment for older community-dwelling people. *BMC Geriatr* 2018;18:103.
- Munro BH. *Statistical methods for health care research*. 4th ed. Lee EH, Chung Y, Kim JS, Song R, Hwang KY, translators. Paju: Koonja Publishing Inc.; 2006.
- Schneider J, Schroth M, Ottenbacher J, Stork W. A novel wearable sensor device for continuous monitoring of cardiac activity during sleep. *Proceedings of the 2018 IEEE Sensors Applications Symposium (SAS)*; 2018 Mar 12-14; Seoul, Korea. Piscataway (NJ): IEEE; 2018.
- Shi J, Mo X, Sun Z. Content validity index in scale development. *Zhong Nan Da Xue Xue Bao Yi Xue Ban* 2012;37:152-5.
- Nunnally JC. An overview of psychological measurement. In: Wolman BB, editor. *Clinical diagnosis of mental disorders: a handbook*. Boston (MA): Springer; 1978. p. 97-146.

18. Raharjanti NW, Wiguna T, Purwadianto A, Soemantri D, Indriatmi W, Poerwandari EK, et al. Translation, validity and reliability of decision style scale in forensic psychiatric setting in Indonesia. *Heliyon* 2022;8:e09810.
19. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. *J Med Internet Res* 2019;21:e16273.
20. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. *JMIR Mhealth Uhealth* 2020;8:e18907.