

Accuracy Bias and Factors Influencing Polysomnography and Consumer Sleep-Monitoring Device Measuring of Total Sleep Time: A Mixed-Methods Study

Jing Yang^{1,2,*}, Dongmei Xu^{3,*}, Huanhuan Lu⁴, Xuwen Yin¹, Haiyan Song¹, Weiwei Zong⁵, Dandan Xu⁵, Xiaohui Lu⁵, Lan Wei⁵, Hong Zhu⁵, Shiyin Zhai⁵, Zejuan Gu¹

¹School of Nursing, Nanjing University of Chinese Medicine, Nanjing, People's Republic of China; ²Nursing Department, The First Affiliated Hospital of Nanjing Medical University, Nanjing, People's Republic of China; ³Rheumatology and Immunology Department, The First Affiliated Hospital of Nanjing Medical University, Nanjing, People's Republic of China; ⁴School of Nursing, Nanjing Medical University, Nanjing, People's Republic of China; ⁵Cardiovascular Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing, People's Republic of China

*These authors contributed equally to this work

Correspondence: Zejuan Gu, School of Nursing, Nanjing University of Chinese Medicine, Nanjing, 210029, People's Republic of China, Email jassicagu@163.com

Objective: The Huawei Band 9 (HWB 9), a consumer sleep-monitoring device with a high market share and a large user base in China, can provide sleep staging parameters and has broad representativeness in sleep monitoring applications. This study aims to compare the accuracy bias of polysomnography (PSG) and consumer sleep-monitoring devices, specifically the HWB 9, in measuring total sleep time (TST) and explore the factors affecting accuracy bias.

Methods: This study employed a sequential explanatory mixed-methods design, with quantitative research comprising 108 samples and qualitative research comprising 18 samples. Select hospitalized patients who required polysomnographic monitoring due to their condition from November 2024 to March 2025 were chosen as the research subjects, and who used PSG and HWB 9 for synchronous sleep monitoring throughout the night. Quantitative data analysis was conducted using descriptive statistics, the Wilcoxon test, Bland-Altman plots, univariate analysis, and multiple linear regression analysis. Qualitative content data were analyzed using NVivo 14.0 software.

Results: The statistical analysis showed a significant difference ($P < 0.05$) between the HWB 9 and PSG in measuring TST. The Bland-Altman plot showed that the measured values deviated from the consistency interval, indicating systematic overestimation bias. The multiple linear regression analysis showed that turning frequency and sleep posture were significant factors affecting measurement bias. Two main themes were found in the qualitative research: sleep habits and environmental factors, and individual differences and psychological perceptions.

Conclusion: Considering the significant variations in individuals, data from such devices should be used with caution in clinical practice.

Keywords: PSG, consumer sleep-monitoring devices, TST, accuracy bias, influencing factors, mixed-methods

Introduction

Sleep is the core physiological process that regulates biological rhythms and plays an irreplaceable role in maintaining normal brain function.¹ In 2015, the Sleep Research Society (SRS) and the American Academy of Sleep Medicine (AASM)² jointly issued a consensus statement stating that adults should ensure at least 7 hours of sleep per day to maintain optimal health. However, with accelerations in the pace of modern life and increases in work pressure, the incidence of sleep disorders continues to rise. Insufficient sleep and the health problems caused by it have become a global public health challenge.³ Total sleep time (TST) is a core indicator for evaluating sleep quality, directly reflecting an individual's sufficient sleep level and serving as an important predictor of physical and mental health. Numerous

studies have shown^{4,5} that insufficient sleep time can lead to various physiological and pathological changes, including impaired immune function, metabolic disorders, emotional disorders, and cognitive decline.⁶ It is also an important risk factor for obesity, diabetes, hypertension, and cardiovascular disease.⁷ In clinical diagnosis and treatment practice, TST is an important basis for diagnosing sleep disorders, such as insomnia and sleep apnea syndrome, and is also a key parameter for evaluating treatment effectiveness. For example, TST is typically significantly decreased in insomnia patients,⁸ while sleep apnea patients⁹ experience a decrease in effective sleep time due to frequent awakenings. Therefore, the accurate monitoring of TST helps to identify sleep disorders early and also provides a scientific basis for developing personalized sleep improvement plans, thereby improving patients' quality of life.

Accurate TST monitoring is the core link in achieving the effective management of sleep disorders. As the "gold standard" for objective sleep monitoring, polysomnography (PSG) can provide accurate sleep staging and related parameter analysis; however, due to limitations such as poor comfort, high technical requirements, and high costs, PSG cannot achieve long-term continuous sleep monitoring.¹⁰ With the development of wearable technology, consumer sleep-monitoring devices using smart bracelets have brought new opportunities for large-scale objective sleep monitoring. This type of device significantly improves user acceptance¹¹ due to its portability, ease of operation, and non-invasiveness. However, TST measurements have systematic biases,^{12–15} and differences in algorithms among different devices lead to inconsistent results,^{16–18} which, to some extent, limit their application value in clinical research and practice. In addition, as the accuracy and available types of sensors continue to increase and core algorithms are constantly upgraded, the accuracy of personal sleep monitoring by consumer sleep-monitoring devices will also continue to improve. This study selected the Huawei Band 9 (HWB 9) as a typical representative consumer sleep-monitoring device. First, this device has a high market share and a large user base in the Chinese consumer market and has a wide range of representativeness in sleep monitoring applications. Second, its True Sleep™4.0 scientific sleep system integrates advanced sleep analysis algorithms and can provide sleep staging parameters. Third, the Huawei Sports and Health App can provide personalized sleep improvement plans and professional insomnia rehabilitation guidance. Other similar devices lack sleep staging functions or have insufficient publicly available information on their algorithms, making it difficult to conduct standardized comparative analysis. A systematic literature review revealed that some high-market share products have been fully validated in previous studies;^{12–15,19} however, no systematic analysis has been conducted on the factors affecting TST measurement bias in such devices. Therefore, this study systematically analyzed the factors affecting measurement bias by comparing PSG and HWB 9 TST measurement results, providing a targeted basis for algorithm optimization, technical improvements, and the clinical application of consumer sleep-monitoring devices.

Methods

Study Design

This study adopted the sequential explanatory mixed-methods design of Creswell and Clarke.²⁰ This method first conducts quantitative research, collects and analyzes the quantitative data, and draws preliminary conclusions. The quantitative research results are used to design and conduct targeted qualitative research to explain, supplement, and explore the quantitative research results in depth.

Participants

In the quantitative research, hospitalized patients at the First Affiliated Hospital of Nanjing Medical University who required polysomnographic monitoring due to their condition from November 2024 to March 2025 were selected as the study subjects. The inclusion criteria were ① an age range of 18–80 years, ② a hospitalization time of ≥ 3 days, and ③ clear consciousness and language expression. The exclusion criteria were ① patients with dementia or other mental illnesses, ② the presence of severe brain organic lesions or other physical illnesses, ③ sleep monitoring for individuals who took hypnotic drugs on the night of sleep, ④ disabled individuals with both upper limbs who are unable to move, and ⑤ patients who were unable to cooperate with this study. After rigorous screening, this study ultimately included 108 cases.

In qualitative research, purposive sampling is used to recruit interviewees from patients participating in the quantitative studies, with high and low TST measurement accuracy used as the standard in this study to ensure the representativeness of the study population. Bertaux²¹ recommends including at least 15 participants in qualitative research to ensure data sufficiency. However, in actual interviews, when newly collected data overlaps with existing data and no new information is provided, the data can be considered to have reached saturation, and interviews are stopped.²² This study included a total of 18 patients.

Data Collection Tools

Data were collected using a general information table designed by researchers based on published literature, PSG, and the HWB 9.

Quantitative Research

Personal Information Form

Personal information form included age, gender, body mass index (BMI), smoking, drinking, hypertension, diabetes, blood type, wrist circumference, skin fold thickness, turning frequency (when the subject changes from any basic position, such as supine, lateral, or prone to another position, such as supine to lateral, it is recorded as one effective turning over. Two uniformly trained research members are on duty in shifts. They continuously observed the subject's position changes and immediately record the type of position change when each turning-over occurs), sleeping position (when the cumulative duration of a certain sleeping position during the subject's nighttime sleep exceeds 40% of the TST,²³ it is the main sleeping position of that night. Sleeping positions are mainly divided into supine, lateral, and prone positions, with the same measurement method as for turning-over times). The severity of obstructive sleep apnea (OSA), the apnea-hypopnea index (AHI) value, awakening frequency, total cholesterol (TC), triglyceride (TG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and lipoprotein levels.

Polysomnography

PSG mainly includes eight electroencephalogram (EEG) leads (F3, F4, C3, C4, O1, O2, M1, and M2) and other physiological parameter acquisition channels. The content of sleep monitoring reports includes monitoring start time, monitoring end time, TST, REM sleep, NREM 1, 2, and 3 stages of sleep, and their respective proportions, awakening frequency, and others.

Consumer Sleep-Monitoring Device

HWB 9 utilizes multimodal biosensing technology and high-precision three-axis accelerometers and photoplethysmography (PPG) to collect real-time vital sign signals, such as heart rate and respiratory rate, from the users. Its True Sleep™4.0 sleep monitoring algorithm can infer wakefulness or sleep stages based on these signals. The content of the sleep monitoring report includes bedtime, wake-up time, TST, and others. Consistent firmware versions were maintained during the study period to ensure algorithm stability.

Qualitative Research

Data were collected using a semi-structured interview outline prepared based on the quantitative analysis results.

Semi-Structured Interview Outline

After systematically reviewing previous literature and clarifying the research objectives, the research team designed interview items based on dual-core requirements: first, to explore the potential behavioral factors that affect the accuracy of consumer sleep-monitoring devices, and second, to explore how user experience affects the measurement results. After the initial draft was generated, the research team conducted two rounds of optimization, first conducting pre-interviews with at least two patients and adjusting the clarity of problem expression based on feedback. Subsequently, clinical sleep medicine experts with PSG qualifications were invited to evaluate the validity of the items from a professional perspective, ultimately creating a structured interview outline ([Supplementary Material 1](#)) to ensure a logical correlation analysis between qualitative data and measurement bias in quantitative research.

Data Collection

In quantitative research, PSG and the HWB 9 were used to monitor the overnight sleep of the study subjects. The specific operation method was as follows: ① On the day of sleep monitoring, drinking beverages that affect sleep, such as strong tea and coffee, was forbidden. ② No other electronic devices or accessories were allowed to be worn except for the research equipment. ③ PSG was installed according to the latest guidelines of the AASM and was professionally analyzed and interpreted. The HWB 9 was worn tightly against the skin on the subject's non-dominant hand at one index finger from the wrist joint. ④ The unified monitoring period was from 19:00 to 07:00 the next day, for a total of 12 hours.

In the qualitative research, face-to-face interviews were conducted with patients with high and low TST measurement accuracy of the HWB 9 in quantitative research. The purpose and significance of the interview were explained to the interviewee before the interview, and the entire interview content was recorded with their consent. Conducting interviews was avoided during the peak periods of ward handover and patient treatment examinations to ensure continuity. During the interview process, the researchers simultaneously recorded changes they observed in nonverbal information, such as facial expressions and eye contact. All of these qualitative data for this study were collected by one of the researchers, who has qualitative research experience. Each interview lasted for 20–30 minutes.

Data Analysis

This study used SPSS 27.0 statistical software to analyze quantitative research data. The categorical variables are expressed as frequency (n) and percentage (%), and continuous variables are expressed as the mean \pm standard deviation. The accuracy bias and overall consistency of the two monitoring methods in measuring TST were compared using the Wilcoxon test and Bland-Altman plots. In the univariate analysis, Pearson (normal distribution) or Spearman (non-normal distribution) correlation coefficients were used for the correlation analysis of continuous variables. Non-parametric tests were used for qualitative data, with the Mann-Whitney *U*-test for binary variables and the Kruskal-Wallis H rank-sum test for multi-categorical variables. Multiple linear regression was used to analyze the factors that affected the accuracy bias of TST measurements. The test level was set at $\alpha = 0.05$.

In the qualitative research, audio recordings were transcribed into transcripts on the day of the interview and cross-checked by two researchers to ensure transcription accuracy. Pseudonyms, such as P1 and P2, were used to identify the interviewees to protect privacy. After the initial draft was completed, the researchers verified the key information with the interviewees to generate the final draft. Before formal analysis, the interviewees were coded in the order of the interviews, and all interview data were imported into NVivo 14.0 software for organization and analysis. Any controversial areas in the analysis were resolved by discussion with the third research member, and the themes were jointly determined.

Data Integration

The triangular design in the mixed method was presented and explained comprehensively, as well as the accuracy bias factors affecting HWB 9 TST measurements from multiple perspectives.

Ethical Considerations

This study was reviewed and approved by the Medical Ethics Committee of the First Affiliated Hospital of Nanjing Medical University (Approval Number: 2024-SR-906). The ethical principles of the Helsinki Declaration were strictly followed during the research process to ensure that the privacy rights and data security of the subjects were fully protected. All participants voluntarily provided written informed consent based on a full understanding of the research content.

Results

Quantitative Results

Accuracy of PSG and HWB 9 TST Measurements

The results in [Table 1](#) indicate a significant difference ($P < 0.05$) between the HWB 9 and PSG in TST measurements, with an average deviation of 40.55 ± 52.41 minutes between the two device measurements. The Bland-Altman plot

Table 1 Results of TST Measurements Using PSG and HWB 9 (n = 108)

Variable	HWB 9		PSG		Bias (Minutes)	Z	p
	Mean±SD (Minutes)	±95% CI (Minutes)	Mean±SD (Minutes)	±95% CI (Minutes)			
TST	458.70±79.33	443.57~473.84	418.16±74.65	403.96~432.40	40.55±52.41	-6.777	<0.001

showed that the measured values deviated from the consistency interval, indicating a systematic overestimation bias and the degree of overestimation was positively correlated with TST ([Supplementary Figure S1](#)).

Sociodemographic Characteristics of the Participants

This study included a total of 115 research subjects, including healthy individuals, patients with OSA, and insomnia patients. After considering missing data, such as data collection failures or technical malfunctions while wearing PSG equipment or the HWB 9, a total of 108 valid samples were obtained ([Table 2](#)). Among them, males accounted for 63%

Table 2 Comparison of Participants' Sociodemographic Characteristics and TST Accuracy Bias

Variables	N=108	Mean ± SD or Percentage	Test	p
Gender			Z=-0.490	0.624
Male	68	63		
Female	40	37		
Age (years)		52.0±11.9	r=0.232	0.016
BMI (kg/m ²)		25.7±3.0	r=0.010	0.915
Wrist circumference(cm)		17.5±1.4	r=-0.140	0.149
AHI		20.2±17.7	r=0.007	0.943
Skinfold thickness (mm)		7.3±2.4	r=-0.057	0.557
Turning frequency (times/h)		4.5±1.5	r=0.589	<0.001
Awakening frequency (times/night)		2.3±1.8	r=0.281	0.003
TC (mmol/L)		4.1±1.0	r=0.199	0.039
TG (mmol/L)		1.5±0.9	r=-0.040	0.679
HDL-C (mmol/L)		1.1±0.2	r=-0.067	0.493
LDL-C (mmol/L)		2.6±0.8	r=0.249	0.009
Lipoprotein (mg/L)		269.8±240.0	r=-0.124	0.201
Sleeping position			Z=-4.455	<0.001
Supine position	50	46.3		
Lateral decubitus position	58	53.7		
Smoking			Z=-2.239	0.025
Yes	24	22.2		
No	84	77.8		
Drinking			Z=-0.470	0.638
Yes	18	16.7		
No	90	83.3		
Hypertension			Z=-0.408	0.683
Yes	60	55.6		
No	48	44.4		
Diabetes			Z=-0.609	0.543
Yes	15	13.9		
No	93	86.1		

(Continued)

Table 2 (Continued).

Variables	N=108	Mean ± SD or Percentage	Test	p
Severity of OSA			H=0.736	0.865
No	20	18.5		
Mild	32	29.6		
Moderate	34	31.5		
Severe	22	20.4		
Blood type			H=6.909	0.075
A	34	31.5		
B	35	32.4		
AB	10	9.2		
O	29	26.9		

Table 3 Multiple Linear Regression Analysis of Factors Affecting the Accuracy Bias of HWB 9 TST Measurements

Independent Variable	Unstandardized Coefficients	Standardized Coefficients		t	p	95% CI	
	B	SE	β			Lower Bound	Upper Bound
Turning frequency (times/h)	12.662	3.835	0.366	3.301	0.001	5.053	20.271
Sleeping position (ref: supine position)							
Lateral decubitus position	26.029	9.625	0.249	2.704	0.008	6.933	45.126

(68 cases), with an age of 52.0 ± 11.9 years. The main nighttime sleeping positions among the research subjects were the supine position (46.3%) and the lateral decubitus position (53.7%), with no preference for prone-position sleeping. The results of the univariate analysis in Table 2 show statistically significant differences ($P < 0.05$) in the accuracy bias of TST measured by the HWB 9 in terms of age, TC levels, LDL-C levels, wakefulness frequency, turning frequency, smoking, and sleeping position.

Factors Affecting the Accuracy Bias of HWB 9 TST Measurements

The age, turning frequency, awakening frequency, TC, and LDL-C were brought in according to the measured values. Smoking and sleeping posture are brought in as categorical variables, and the assignment is as follows. Smoking: yes =0 (reference group), no =1. Sleeping position: supine position =0 (reference group), lateral decubitus position=1. The results of multiple regression analysis showed statistical significance (adjusted $R^2 = 0.255$, $F = 6.230$, $P < 0.001$). As shown in Table 3, turning frequency and lateral decubitus position were found to be statistically significant factors affecting accuracy bias ($P < 0.05$).

Qualitative Results

Eighteen patients were included in the qualitative research (Table 4), with an average age of 50 years (24–64 years). Of them, 61% were male, and 89% were married.

Theme 1: Sleep Habits and Environmental Factors

The qualitative research showed that sleep habits and environmental factors affected the accuracy of HWB 9 measurements through two pathways: device displacement and the misjudgment of sleep status.

In terms of device displacement, qualitative interviews and nighttime dynamic angle monitoring showed that measurement accuracy was better when the device was horizontally compressed (ie, at a perpendicular angle to the arm’s long axis), whereas deviations were more likely to occur when the device was parallel to the arm’s long axis.

Table 4 Characteristics of the Participants in the Qualitative Phase

Participant (P)	Age	Gender	Education Level	Marital Status	TST Accuracy Bias (Minutes)
P1	61	Male	Senior middle school	Married	High (0.4)
P2	50	Male	Junior college	Married	Low (111.8)
P3	59	Female	Middle school	Married	High (4.3)
P4	57	Female	Middle school	Married	Low (69.5)
P5	31	Male	Undergraduate	Married	Low (58.2)
P6	49	Female	Middle school	Married	High (-4.4)
P7	61	Male	Primary school	Married	High (-7.5)
P8	58	Male	Middle school	Married	Low (64.8)
P9	24	Female	Postgraduate	Single	High (3.7)
P10	57	Male	Middle school	Married	Low (87)
P11	30	Male	Undergraduate	Single	High (7.1)
P12	59	Female	Primary school	Married	Low (106.4)
P13	64	Male	Middle school	Married	Low (77.7)
P14	33	Male	Junior college	Married	High (-5.1)
P15	45	Male	Senior middle school	Married	High (4.6)
P16	51	Female	Middle school	Married	High (7.7)
P17	61	Male	Middle school	Married	Low (97.8)
P18	58	Female	Middle school	Married	Low (112.1)

When I can't sleep at night, I like to frequently turn over and adjust my sleeping position. (P2)

I like to sleep on my side most of the time at night, switching back and forth. (P17)

In terms of misjudging sleep state, ambient light may prolong the wakefulness period by inhibiting melatonin secretion, while the device fails to capture the intermediate state of closed-eye rest but actual wakefulness, resulting in a bias in recognizing the wakefulness-sleep transition threshold.

I have shallow sleep, and when there is a little noise, I will wake up and not be able to sleep, but I still close my eyes and pretend to sleep. (P18)

I am used to turning on the night light when I sleep, which is convenient for getting up and going to the bathroom. (P10)

In addition, patients' preference for pillow height may indirectly affect the quality of signal acquisition by changing the relative position of the body and the device.

When I sleep at night, I am used to using a higher pillow, which feels more comfortable (P8)

Due to cervical issues, I do not use a pillow at night. (P12)

Theme 2: Individual Differences and Psychological Perception

Individual physiological characteristics and psychological states exacerbate measurement errors from two dimensions: signal acquisition efficiency and subjective cognitive bias.

At the individual physiological level, a patient's repeated position adjustment at night, excessive or thin wrist circumference, which causes unstable device adhesion, and poor sensor contact due to skin sweating may interfere with the continuity of PPG signals.

I feel that the material of the wristband is uncomfortable, and I have to adjust its position repeatedly at night. (P13)

My wrist circumference is too narrow, and even when it is pressed to the smallest position, it still feels loose. (P12)

My wrist circumference is a bit thick. At first, I felt the tightness was moderate when I put it on, but halfway through sleep, it felt a bit tight. So I adjusted it a little by myself. (P2)

Table 5 Integrated Summary of Quantitative and Qualitative Factors Affecting the Accuracy Bias of HWB 9 TST Measurements

Theme	Quantitative Evidence	Qualitative Insight
Validated behavioral patterns	A significantly statistical difference was observed in the accuracy of the number of flips ($P = .001$). When lying on the side, the difference in the accuracy of TST measured by the HWB 9 was better than that in the supine position ($P = 0.008$).	Frequent turning over at night and adjusting sleeping posture or lying on one's side can cause compression of the arms, leading to the amplification of motion artifacts, sensor displacement, and signal obstruction, resulting in data misinterpretation.
Factors related to individual differences	Age ($r = 0.232$, $p = 0.016$), TC levels ($r = 0.199$, $p = 0.039$), and LDL-C levels ($r = 0.249$, $p = 0.009$) were positively correlated with differences in accuracy.	Elderly patients have more fragmented nighttime sleep, and insufficient device recognition leads to increased bias. If a patient's nighttime sweat secretion increases or the skin produces oil, adhesion between the device and the skin is decreased, which may interfere with the quality of PPG signal acquisition.
Supplemental psychological and environmental factors		Patients are concerned about missing sleep data, which may lead to an increase in nighttime wakefulness and indirectly widen measurement differences. Patients may experience fragmented sleep or difficulty falling asleep due to strong light stimulation at night, leading to the poor ability of the device to recognize micro-arousal states and resulting in increased measurement bias.

When I sleep at night, the place where I wear the device is prone to sweating. (P17)

At the psychological level, patients' excessive concerns about the accuracy of device measurements and the psychological burden of wearing them can lead to psychological stress, which may result in a cumulative effect of subjective awakening and recognition defects in device algorithms.

When I wear the device at night, I feel a bit mentally burdened, especially when I can't sleep, I feel anxious and worried about inaccurate measurements (P4, 18)."

I feel like I didn't sleep well last night and didn't get enough sleep in total. (P5)

But discrepancies between objective measurements suggest that patients may underestimate their actual sleep time, leading to discrepancies between subjective and objective sleep assessments, which are consistent with Stephan and Siclari's²⁴ results.

Integration Results of Mixed Research

The quantitative and qualitative research results were integrated. Table 5 presents the validated behavioral patterns, factors related to individual differences, and supplemental psychological and environmental factors involved in accuracy bias in HWB 9 TST measurements.

Discussion

With the development of the social economy and advances in medical technology, people's attention to sleep health is increasing, and their understanding of sleep disorders is also deepening. The AASM emphasizes the importance of long-term management and effective evaluations of sleep disorders.²⁵ In recent years, with continuous innovations in sensing technology and artificial intelligence algorithms, consumer sleep-monitoring devices, such as smart bracelets, have significantly improved their performance.²⁶ Whether they can be used as a substitute for PSG has aroused strong interest among researchers and clinicians. However, before a consumer sleep-monitoring device is applied in clinical practice, it needs to be compared with PSG to confirm whether it can provide data support for correct clinical decision-making.²⁷ Therefore, this study used PSG as the gold standard to systematically evaluate the accuracy bias and factors affecting HWB 9 TST measurements.

This was a mixed-methods study, providing statistical evidence for bias phenomena using quantitative results and revealing its influencing factors using qualitative data. The quantitative results showed significant differences between

HWB 9 and PSG TST measurements, with an average deviation of 40.55 minutes ($P < 0.05$). Although the error range was within the clinically acceptable range, the analysis found that the measurement deviation was particularly prominent in the elderly population due to the high degree of sleep fragmentation, which is consistent with research by Danzig et al.²⁸ At the same time, research found that the HWB 9 poorly recognizes the transition between wakefulness and sleep states. The more awake patients are at night and the longer they stay awake, the greater the measurement errors of the device, consistent with the results of Hamill et al.^{8,14,29,30} On this basis, qualitative research serves as a supplement to quantitative research, fully revealing the pathway of sleep fragmentation → device recognition defects → bias amplification, filling the mechanism gap in statistical models. In addition, the quantitative analysis found that HWB 9 TST measurements were better in healthy individuals than those with sleep problems, which is consistent with the findings of Moscoso et al.¹⁹ Qualitative interviews further elucidated the mechanism by which individual differences may affect accuracy bias by providing qualitative descriptions of anxiety and patients' subjective insomnia but overestimating or underestimating the duration of equipment use and signal loss caused by wrist discomfort.

The deep integration of two types of data presents a triple complementary effect, significantly improving the robustness of research conclusions. Qualitative research provides in-depth explanations for quantitative results from three dimensions: user behavior, physiological experience, and psychological perception. At the user behavior level, qualitative interviews provided feedback on frequently turning over to adjust sleeping posture at night when sleep quality is poor, wristband displacement caused by arm pressure when lying on the side, and active position adjustment due to the discomfort of the wristband at night. This confirmed that the turning-over times and lying on the side were statistically significant factors ($P < 0.05$) affecting accuracy differences, revealing a causal chain of position changes → signal occlusion → data misjudgment, which is consistent with the conclusion of Liang and Chapa-Martell³¹ that body movements may produce motion artifacts that interfere with signal acquisition. Further research found that frequent turning over is a behavioral manifestation of sleep discomfort, which amplifies measurement bias through the dual effects of increasing the number of awakenings and blind spots in device signal acquisition. At the physiological level, the quantitative analysis showed a positive correlation between TC levels, LDL-C levels, and measurement bias. Qualitative interviews indicated that skin oil or sweat secretion could lead to poor device contact, indicating that abnormal blood lipid levels may indirectly interfere with PPG signal quality. At the psychological level, the qualitative results showed that patients were worried about missing sleep data but instead became nervous and unable to sleep all night, revealing that psychological anxiety triggers micro-awakenings, indirectly expanding measurement bias and adding the involvement of psychological factors to the mechanism not covered by quantitative analysis. In addition, the qualitative research captured environmental factors such as night lights, further expanding the dimensions of environmental psychological interactions. By integrating the quantitative identification of significant factors, the qualitative revelation of action pathways, and performing cross-validation to improve theory, a four-in-one bias explanation framework of behavior-physiology- psychology-environment was constructed, which significantly improved the robustness and clinical application value of the research conclusions.

This study has limitations in methodology, sample selection, and the experimental environment, which may affect the accuracy of measurement bias assessment and the generalizability of the conclusions. First, at the methodological level, single-night sleep monitoring cannot capture the diurnal or weekly fluctuations of individual sleep patterns and lacks cross-validation from multi-night data, making it difficult to distinguish the interference caused by inherent device errors and sleep state fluctuations, resulting in the fortuity of bias measurement results. Second, there was bias in sample selection. Due to considerations of data integrity and nighttime behavior observations, the study only included hospitalized patients, which may have resulted in sample selection bias. Considering the inherent heterogeneity of OSA, insomnia, and healthy populations and the fact that hospitalized patients may be influenced by disease, treatment, and environmental factors, their sleep behavior and physiological responses may differ from those of individuals at home, which limits the extrapolation of conclusions to the general population. Once again, at the experimental environment level, although measurements were taken 3 days after the patient's hospitalization to alleviate environmental adaptation issues, PSG monitoring required maintaining a static operating standard, which changed the participants' natural sleep patterns and affected the accuracy of HWB 9's algorithm recognition based on normal sleep states. In addition, the equipment in the experiment was under ideal conditions of known sleep periods and did not need to autonomously

determine the time of falling asleep and waking up. However, in real-world use, users have random sleep patterns and unpredictable nighttime activities, and the equipment needs to independently perform TST recognition. This difference may have led to the underestimation of the inaccuracy of the equipment under daily conditions in this study. Finally, regarding the issue of device adaptability, the study only focused on the HWB 9 as a single device, and its algorithm characteristics and sensor configurations do not represent the technological diversity of all consumer sleep devices.

Although consumer sleep-monitoring devices have technological limitations in TST measurements, their potential for application in the field of health management is enormous. These devices provide users with dynamic health data by continuously monitoring core indicators, such as sleep duration and sleep structure, promoting the transformation of medical services from passive diagnosis and treatment to active management, helping chronic disease prevention and control strategies move from disease treatment to health maintenance, and providing technical support for constructing a national preventive health management system. However, achieving their comprehensive application requires addressing three key challenges: in terms of technological optimization, quantifying the impact of sleep habits, environmental factors, and individual differences on measurement accuracy, developing targeted calibration algorithms for the elderly population, and optimizing device-wearing stability and sensor sensitivity. At the level of improving the user experience, standardized wearing guidelines should be developed while also considering users' physiological comfort and psychological acceptance to enhance device usage compliance. At the clinical application level, a sleep health warning system should be constructed based on user sleep behavior characteristics to cover management over the entire lifecycle and explore remote monitoring and intervention modes. In the future, with technological innovations and the continuous optimization of user experiences, consumer sleep-monitoring devices are expected to become an important tool for personal health management.

Conclusion

This study adopted a sequential explanatory mixed-methods design, which revealed statistical patterns using quantitative data and explained the mechanism of action using qualitative analysis, forming a methodological triangle for mutual verification. The accuracy bias of PSG and consumer sleep-monitoring devices represented by the HWB 9 in measuring TST was assessed, and the factors that affected measurement bias were explored. Using PSG as the gold standard, the research results found that the HWB 9 has significant overestimation bias, especially in the elderly population, and performs better in measuring TST in healthy individuals than in individuals with sleep problems. According to observations, sleep habits, individual differences, and environmental and psychological perceptions are important factors contributing to accuracy bias. The conclusion drawn is that greater accuracy bias may occur in patients who turn frequently at night and maintain a lateral sleeping position. Considering the significant variations in individuals, data from such devices should be used with caution in clinical practice.

Data Sharing Statement

The datasets generated during and/or analyzed during the current study are not publicly available, because some of the data involves personal privacy, but are available from the corresponding author on reasonable request.

Consent to Publish

This study has obtained consent to publish from each study participant.

Author Contributions

Jing Yang: Conceptualization, Data curation, Formal analysis, Investigation, Resources, Software, Writing-original draft, Writing-review & editing. Dongmei Xu: Conceptualization, Formal analysis, Investigation, Resources, Supervision, Writing-review & editing. Huanhuan Lu: Conceptualization, Methodology, Software, Writing-original draft. Xuwen Yin: Conceptualization, Data curation, Investigation, Methodology, Writing-original draft. Haiyan Song: Data curation, Investigation, Supervision, Writing-original draft. Dandan Xu: Formal analysis, Investigation, Supervision, Writing-review & editing. Weiwei Zong: Data curation, Formal analysis, Supervision, Writing-review & editing. Xiaohui Lu: Formal analysis, Supervision, Validation, Writing-review & editing. Lan Wei: Investigation, Methodology, Supervision,

Writing–review & editing. Hong Zhu: Conceptualization, Data curation, Formal analysis, Writing–review & editing. Shiyin Zhai: Conceptualization, Data curation, Visualization, Writing–review & editing. Zejuan Gu: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing–original draft, Writing – review & editing.

All authors gave final approval of the version to be published, have agreed on the journal to which the article has been submitted, and agreed to be accountable for all aspects of the work.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Jiangsu Province Postgraduate Practice Innovation Project [grant number SJCX25_0915]; Jiangsu Province Hospital [grant number CXTDA2017019].

Disclosure

All authors declare no competing interests.

References

1. Matenchuk BA, Mandhane PJ, Kozyrskyj AL. Sleep, circadian rhythm, and gut microbiota. *Sleep Med Rev*. 2020;53:101340. doi:10.1016/j.smrv.2020.101340
2. Watson NF, Badr MS, Belenky G, et al. Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society. *Sleep*. 2015;38(6):843–844. doi:10.5665/sleep.4716
3. Sateia MJ. International classification of sleep disorders-third edition: highlights and modifications. *Chest*. 2014;146(5):1387–1394. doi:10.1378/chest.14-0970
4. Bain AR, Weil BR, Diehl KJ, et al. Insufficient sleep is associated with impaired nitric oxide-mediated endothelium-dependent vasodilation. *Atherosclerosis*. 2017;265:41–46. doi:10.1016/j.atherosclerosis.2017.08.001
5. Namtvedt SK, Hisdal J, Randby A, et al. Impaired endothelial function in persons with obstructive sleep apnoea: impact of obesity. *Heart*. 2013;99(1):30–34. doi:10.1136/heartjnl-2012-303009
6. Yu Z, Bingqian Z, Ni G, et al. Dual trajectory of sleep and frail in elderly people. *J Central South Univ*. 2023;48(4):621–627. doi:10.11817/j.issn.1672-7347.2023.220544
7. Wilson S, Anderson K, Baldwin D, et al. British Association for Psychopharmacology consensus statement on evidence-based treatment of insomnia, parasomnias and circadian rhythm disorders: an update. *J Psychopharmacol*. 2019;33(8):923–947. doi:10.1177/0269881119855343
8. Dong X, Yang S, Guo Y, et al. Validation of Fitbit Charge 4 for assessing sleep in Chinese patients with chronic insomnia: a comparison against polysomnography and actigraphy. *PLOS ONE*. 2022;17(10):e0275287. doi:10.1371/journal.pone.0275287
9. Moreno-Pino F, Porras-Segovia A, López-Esteban P, et al. Validation of Fitbit Charge 2 and Fitbit Alta HR against polysomnography for assessing sleep in adults with obstructive sleep apnea. *J Clin Sleep Med*. 2019;15(11):1645–1653. doi:10.5664/jcsm.8032
10. Rundo JV, Downey R. Polysomnography. *Handbook Clin Neurol*. 2019;160:381–392. doi:10.1016/b978-0-444-64032-1.00025-4
11. Park KS, Choi SH. Smart technologies toward sleep monitoring at home. *Biomed Eng Lett*. 2019;9(1):73–85. doi:10.1007/s13534-018-0091-2
12. Chinoy ED, Cuellar JA, Huwa KE, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep*. 2021;44(5). doi:10.1093/sleep/zsaa291
13. Frijia J, Mullaert J, Abensur Vuillaume L, et al. Metrology of two wearable sleep trackers against polysomnography in patients with sleep complaints. *J Sleep Res*. 2025;34(2):e14235. doi:10.1111/jsr.14235
14. Haghayegh S, Khoshnevis S, Smolensky MH, et al. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. *J Med Internet Res*. 2019;21(11):e16273. doi:10.2196/16273
15. Schyvens AM, Van Oost NC, Aerts JM, et al. Accuracy of Fitbit Charge 4, Garmin Vivosmart 4, and WHOOP versus polysomnography: systematic review. *JMIR mHealth uHealth*. 2024;12:e52192. doi:10.2196/52192
16. de Zambotti M, Goldstone A, Claudatos S, et al. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. *Chronobiol Int*. 2018;35(4):465–476. doi:10.1080/07420528.2017.1413578
17. Lim SE, Kim HS, Lee SW, et al. Validation of Fitbit Inspire 2™ against polysomnography in adults considering adaptation for use. *Nat Sci Sleep*. 2023;15:59–67. doi:10.2147/nss.S391802
18. Robbins R, Weaver MD, Sullivan JP, et al. Accuracy of three commercial wearable devices for sleep tracking in healthy adults. *Sensors*. 2024;24(20):6532. doi:10.3390/s24206532
19. Concheiro-Moscoso P, Groba B, Alvarez-Estevéz D, et al. Quality of sleep data validation from the Xiaomi Mi Band 5 against polysomnography: comparison study. *J Med Internet Res*. 2023;25:e42073. doi:10.2196/42073
20. Creswell JW, Creswell JD. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage; 2017.
21. Bertaux D. *Biography and Society: The Life History Approach in the Social Sciences*. Beverly Hills, California: SAGE Publications; 1981.
22. Grady MP. *Qualitative and Action Research: A Practitioner Handbook*. Bloomington, Indiana: Phi Delta Kappa Educational Foundation; 1998.
23. Zhang Y, Xiao A, Zheng T, et al. The relationship between sleeping position and sleep quality: a flexible sensor-based study. *Sensors*. 2022;22(16). doi:10.3390/s22166220
24. Stephan AM, Siclari F. Reconsidering sleep perception in insomnia: from misperception to mismeasurement. *J Sleep Res*. 2023;32(6):e14028. doi:10.1111/jsr.14028
25. Singh J, Badr MS, Diebert W, et al. American Academy of Sleep Medicine (AASM) position paper for the use of telemedicine for the diagnosis and treatment of sleep disorders. *J Clin Sleep Med*. 2015;11(10):1187–1198. doi:10.5664/jcsm.5098

26. Depner CM, Cheng PC, Devine JK, et al. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. *Sleep*. 2020;43(2). doi:10.1093/sleep/zsz254
27. Menghini L, Cellini N, Goldstone A, et al. A standardized framework for testing the performance of sleep-tracking technology: step-by-step guidelines and open-source code. *Sleep*. 2021;44(2). doi:10.1093/sleep/zsaa170
28. Danzig R, Wang M, Shah A, et al. The wrist is not the brain: estimation of sleep by clinical and consumer wearable actigraphy devices is impacted by multiple patient- and device-specific factors. *J Sleep Res*. 2020;29(1):e12926. doi:10.1111/jsr.12926
29. Chase JD, Busa MA, Staudenmayer JW, et al. Sleep measurement using wrist-worn accelerometer data compared with polysomnography. *Sensors*. 2022;22(13):5041. doi:10.3390/s22135041
30. Hamill K, Jumabhoy R, Kahawage P, et al. Validity, potential clinical utility and comparison of a consumer activity tracker and a research-grade activity tracker in insomnia disorder II: outside the laboratory. *J Sleep Res*. 2020;29(1):e12944. doi:10.1111/jsr.12944
31. Liang Z, Chapa-Martell MA. Accuracy of Fitbit wristbands in measuring sleep stage transitions and the effect of user-specific factors. *JMIR mHealth uHealth*. 2019;7(6):e13384. doi:10.2196/13384

Nature and Science of Sleep

Publish your work in this journal

Nature and Science of Sleep is an international, peer-reviewed, open access journal covering all aspects of sleep science and sleep medicine, including the neurophysiology and functions of sleep, the genetics of sleep, sleep and society, biological rhythms, dreaming, sleep disorders and therapy, and strategies to optimize healthy sleep. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/nature-and-science-of-sleep-journal>

Dovepress
Taylor & Francis Group