BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Reliability and validity of a novel single-lead portable electrocardiogram device for pregnant women: a comparative study

Haixue Wang[1†], Jianwei Wang[2†], Wei Jing[2], Shanshan Dai[2], Deyun Zhang[3], Shijia Geng[3], Haijun Wang[4*] and Shenda Hong[5,6,7*]

Abstract

Background WenXinWuYang, a novel portable Artificial Intelligence Electrocardiogram (AI-ECG) device, can detect many kinds of abnormal heart disease and perform a single-lead ECG, but its reliability and validity among pregnant women is unclear. The aim of this study was to assess the reliability and validity of heart rate, ECG measurements and diagnostic results by compared the portable device with a clinical 12-lead ECG among pregnant women.

Methods We conducted a clinical study at a municipal-level maternal and child health care hospital. The pregnant women who visited the ECG room for ECG examination were invited to participate in this study. Each participant underwent three ECG recordings: one with conventional 12-lead ECG and two with WenXinWuYang ECG. The first WenXinWuYang ECG was recorded simultaneously with the 12-lead ECG. We collected heart rate, ECG measurements, and diagnostic results related to arrhythmias from both ECG devices. The data were then analyzed using Spearman rank correlation coefficients, consistency analyses, and Bland-Altman plots.

Results The study included 287 ECG recordings from 99 pregnant women, with a balanced distribution across different stages of pregnancy. We observed strong to moderate correlations between the two WenXinWuYang measurements for heart rate ($r=0.847$), PR interval ($r=0.728$), QRS duration ($r=0.636$), QT interval ($r=0.836$), and QTc interval ($r=0.648$), with a diagnostic consistency rate exceeding 90.0%. When compared with the 12-lead ECG, the mean differences for heart rate, PR interval, QRS duration, QT interval, and QTc interval were -0.4 ± 3.1 bpm ($r=0.957$), 14.6 ± 12.4 ms ($r=0.537$), 7.0 ± 8.9 ms ($r=0.136$), 2.1 ± 12.0 ($r=0.774$), and 6.6 ± 16.5 ($r=0.663$), respectively. Although the correlation coefficient was low in QRS duration, Bland-Altman results showed moderate to strong agreement between these intervals. Sinus rhythm recognition was fully consistent with the 12-lead ECG, with higher validity in detecting arrhythmias (sensitivity 84.2%, specificity 97.5%). Similar trends existed among different stages of pregnancy.

[†]Haixue Wang and Jianwei Wang contributed equally to this work.

*Correspondence:
Haijun Wang
whjun1@bjmu.edu.cn
Shenda Hong
hongshenda@pku.edu.cn

Full list of author information is available at the end of the article

Conclusion WenXinWuYang demonstrated acceptable reliability and validity in measuring heart rate, ECG measurements, and detecting arrhythmias among pregnant women.

Keywords Portable device, ECG, Reliability, Validity, 12-lead ECG

Background

Cardiovascular diseases (CVDs) have become the leading cause of age-standardized deaths worldwide [1], with mortality from CVDs continuing to rise in many regions [2] and expected to grow to more than 23.6 million by 2030 [3]. As a non-invasive and cost-effective detecting tool, the electrocardiogram (ECG) plays an important role in the early detection, monitoring, and management of various heart conditions, contributing to the prevention and treatment of cardiovascular diseases.

The advent of the portable ECG devices holds great promise for revolutionizing how healthcare professionals and individuals monitor cardiac condition. These devices not only address the challenges of infrequent hospital visits, but also enable the timely detection of heart abnormalities [4]. Portable ECG devices offer several advantages, including their small size, wireless capabilities, real-time monitoring, and ease of use, making them suitable for monitoring heart rhythms and waveforms [5, 6]. Although some portable ECG devices have been proven accurate or promising for atrial fibrillation screening [7–10], heart rate variability (HRV) assessment [11], and ECG measurements in real-world settings [12, 13], many of them offer limited information compared to traditional 12-lead ECGs [14]. Furthermore, evidence about their consistency and accuracy in real-world scenarios remains limited and unsatisfactory [5, 6, 11, 15–17]. In addition, a significant portion of this evidence originates from high-income countries, with only a limited number of studies conducted in low- and middle-income countries, including China [5, 15].

High-quality instruments are useful tools for both research and clinical purposes [18]. Reliability and validity are among the most important and fundamental aspects for accessing the accuracy of any measuring methodology used in rigorous research [19]. Reliability, or consistency, refers to the stability of a measurement instrument, indicating how consistently it produces the same results on separate occasions (such as the commonly used test–retest reliability) [19]. Validity is the degree to which an instrument measures what it is intended to measure. Criterion-related validity is one of the common indicators for evaluating the validity, which was conducted by comparing the instrument with a criterion measure that has been established as valid (gold standard) [19, 20]. To determine whether an instrument has high quality, measurement properties such as reliability and validity need to be assessed by standardized criteria. Given the limited evidence on the accuracy of portable ECG devices, it is of great significance to conduct high-quality evaluation studies using standardized method.

China's rapid economic growth has been accompanied by a sharp increase in cardiovascular disease [21], with an estimated 330 million individuals suffering from CVDs [22]. This has brought a significant economic burden to society, hospitals and families. In response to this challenge and the insufficient of portable ECG devices in China, WenXinWuYang was developed using a deep learning model trained on millions of real ECG data samples [23, 24]. This portable device offers professional-grade single-lead ECG monitoring and is certified for accuracy and reliability by the National Medical Products Administration of China (NMPA) (Anhui Province Medical Device Registration No. 20202070406). Users can easily connect the device to their smartphones via Bluetooth, requiring no complex setup, and monitor their health status anytime and anywhere [25]. Although the device has shown strong performance in detecting nearly all arrhythmias based on an external dataset test [23], its effectiveness in real-world clinical practice remains unknown.

Women confronted with an additional burden of gender-specific risk factors, excepted for traditional risk factors associated with cardiovascular disease. Key stages of a woman's reproductive history may influence or reveal short- and long-term cardiometabolic and cardiovascular trajectories [26]. Pregnancy is a special period which is easily to be affected by various types of heart disease due to the increasing heart load with the increase of gestational age. Early cardiac monitoring for pregnant women is essential in clinical practice, but it only targeted at high-risk individuals in the middle and late pregnant stages [27]. On one hand, this might overlook some patients [28], on the other hand, high-risk pregnant women need to visit hospitals for 12-lead or invasive examinations [29]. In addition, the duration of routine prenatal check-ups for pregnant women does not allow for continuous cardiac health monitoring, but frequent visits to the hospital are inconvenient and also increase the medical burden. Therefore, it is necessary to monitor pregnant women's cardiac health by using portable ECG devices. However, it's not clear how does this portable ECG device perform among pregnant women. Thus, we performed a real-world study in a maternal and child health care hospital among pregnant women to assess the reliability and validity of WenXinWuYang in monitoring

ECG measurements and detecting of arrhythmias by comparing it with 12-lead ECG.

Methods

To evaluate the reliability and validity of WenXinWuYang among pregnant women by comparing with 12-lead ECGs, this study was conducted in Weifang Maternal and Child Health Care Hospital, Shandong province, China, from June 2023 to January 2024. Pregnant women who went to the ECG room for a routine visit testing were invited to participate in this study if they met the following inclusion criteria: aged between 18 and 45 years, diagnosed with pregnancy. Exclusion criteria included refused to provide informed consent, with a history of hypertension, diabetes, thyroid dysfunction, polycystic ovary syndrome, respiratory diseases, epilepsy or psychiatric disorders, taken steroid drugs or psychotherapeutic drugs, with a history of smoking, drinking or substance abuse. Considering pregnancy is a special period, to better assessing the effectiveness of detection in pregnant women, we took the balanced distribution of women's pregnant stage into account. All participants would be fully informed detail information regarding the study and written consent was obtained before data collection. Ethical approval was granted by the Medical Ethics Committee of Weifang Maternal and Child Health Hospital (Number 2023 – 211).

We used test-retest reliability to evaluate the reliability of this portable ECG device, which meant that women will be measured twice by WenXinWuYang, and evaluated its validity by comparing with the golden 12-lead ECG (Electrocardiograph, ECG-2360, Japan Photoelectric Industry Co., LTD, Japan). Before collecting the ECG data, participants' age, weight, height, gestational week, parity of pregnancy would be asked, then each pregnant woman was measured by WenXinWuYang twice. The first measurement was performed simultaneously with the clinical 12-lead ECG, with the participant holding the portable ECG device using both thumbs and index fingers (Figure S1 in Supplement). Participants were asked to lie and keep still during the 30-second test. The second measurement was conducted immediately (within 60 s) after the 12-lead ECG, with the participant again holding the single-lead ECG for another 30 s while lying still. During both measurements, we monitored the signal quality of WenXinWuYang to ensure the quality of the data. If any human-caused signal interference appeared, or if a continuous stable waveform was recorded for less than 15 s, the participant was retested with WenXinWuYang. If poor signal quality was observed during the first measurement, the 12-lead ECG was also retested alongside WenXinWuYang. The 12-lead ECG results were directly printed, and the portable ECG results were automatically saved in a Wechat mini program on a smartphone, then de-identified and downloaded for further analysis.

Heart rate, ECG measurements including PR interval, QRS duration, QT interval, QTc interval (calculated using the Bazett formula), and diagnostic results collected by both ECG devices were recorded for analyses. Considering the portable single-lead ECG has its limitation in detecting myocardial ischemia, to ensure the comparability of single-lead and 12-lead measurement results, only heart-rhythm related results were compared. Test-retest reliability was used to evaluate the stability of portable ECG measurement results. The validity of the portable device was evaluated by comparing it with 12-lead ECG.

Statistical analyses

Continuous variables with normal distribution were presented as mean ± standard deviation, continuous variables with abnormal distribution were presented as medians with interquartile ranges (IQR). Categorical variables were described as numbers and percentages (%). Basic characteristics were compared using Kruskal-Wallis rank sum test for continuous variables with abnormal distribution and Fisher's exact test for categorical variables. Spearman correlation was used for continuous variables with abnormal distribution, and Pearson correlation was used for continuous variables with normal distribution between WenXinWuYang and 12-lead ECG. Coefficient $r > 0.70$ was considered as a strong correlation, while $r > 0.60$ was considered as a moderate correlation, with a stronger correlation indicating greater consistency. Bland-Altman analysis was used to assess the level of agreement between devices. Strong agreement was defined as interval differences of less than 20 ms and heart rate differences of less than 5 bpm, while moderate agreement was defined as interval difference of less than 40 ms [30].

The diagnostic performance of WenXinWuYang was assessed using accuracy, sensitivity, and specificity, which were calculated according to Eqs. (1)-(3) [31, 32]. In these equations, TP and TN indicate the number of positive and negative samples that are correctly detected by WenXinWuYang. FP specifies the number of negative samples incorrectly predicted as positive, while FN is the number of positive samples incorrectly predicted as negative by WenXinWuYang [33]. The diagnostic consistency rate of test-retest reliability was also calculated using Eq. (1). The results were considered statistically significant at two-sided $p < 0.05$. Statistical analyses were performed using R software (version 4.3.0; creator: John Chambers and colleagues; location: Jersey City, NJ, USA).

$$Accuracy = \frac{\mathrm{T}P + TN}{\mathrm{T}P + TN + FP + FN} \quad (1)$$

Table 1 The characteristics of enrolled pregnant women ($N=99$)[a]

Characteristics	Overall	First trimester ($n=34$)	Second trimester ($n=32$)	Third trimester ($n=33$)	*p*
Age(years)	29.0 (27.0, 33.0)	28.5 (27.0, 33.8)	30.0 (27.8, 34.0)	30.0 (27.0, 32.0)	0.701[b]
Height(cm)	163.7 (6.0)	162.6 (5.3)	163.0 (5.8)	165.4(6.8)	0.104[b]
Weight(kg)	65.0 (59.0, 74.5)	61.0 (56.0, 68.0)	65.0 (57.8, 71.0)	73.0 (63.0, 78.0)	0.002[b]
Parity					0.874[c]
Firstborn	59 (59.6)	22 (64.7)	17 (53.1)	20 (60.6)	
Second child	29 (29.3)	8 (23.5)	11 (34.4)	10 (30.3)	
Third child	11 (11.1)	4 (11.8)	4 (12.5)	3 (9.1)	

[a] Categorical variables were reported as n(%) and continuous variables with normal distribution were reported as mean(SD), continuous variables with abnormal distribution were reported as medians (IQR)

[b] Kruskal-Wallis rank sum test

[c] Fisher's exact test

Table 2 The continuous variables test-retest reliability of WenXinWuYang [a]

Parameters	Parameters	WenXinWuYang first time	WenXinWuYang second time	r[b]	*p*
Total	HR	81.0 (75.0, 91.0)	83.0 (76.0, 91.5)	0.847	< 0.001
	PR interval	152.0 (145.0, 160.0)	152.0 (146.0, 159.0)	0.728	< 0.001
	QRS duration	96.0 (89.5, 98.5)	96.0 (89.0, 96.0)	0.636	< 0.001
	QT interval	374.0 (360.5, 381.0)	374.1 (15.0)	0.836	< 0.001
	QTc interval	438.3 (22.4)	443.3 (23.9)	0.648	< 0.001
First trimester	HR	82.5 (78.0, 90.5)	83.5 (76.8, 90.8)	0.897	< 0.001
	PR interval	152.0 (145.0, 160.0)	155.5 (146.8, 160.5)	0.569	< 0.001
	QRS duration	96.0 (89.0, 100.3)	92.0 (89.0, 96.0)	0.458	0.006
	QT interval	373.5 (361.3, 379.8)	373.3 (14.7)	0.793	< 0.001
	QTc interval	438.5 (23.0)	437.6 (22.3)	0.609	< 0.001
Second trimester	HR	78.0 (72.0, 85.3)	78.5 (75.0, 89.0)	0.776	< 0.001
	PR interval	152.5 (146.8, 160.3)	157.0 (151.8, 159.3)	0.741	< 0.001
	QRS duration	96.0 (92.0, 96.0)	96.0 (89.0, 96.0)	0.759	< 0.001
	QT interval	377.5 (358.8, 393.0)	378.6 (15.7)	0.813	< 0.001
	QTc interval	435.6 (21.2)	439.0 (21.5)	0.632	< 0.001
Third trimester	HR	82.0 (77.0, 95.0)	87.0 (79.0, 97.0)	0.807	< 0.001
	PR interval	147.0 (144.0, 158.0)	147.0 (144.0, 153.0)	0.833	< 0.001
	QRS duration	96.0 (90.0, 99.0)	96.0 (89.0, 96.0)	0.732	< 0.001
	QT interval	369.0 (362.0, 380.0)	370.6 (14.1)	0.910	< 0.001
	QTc interval	440.6 (23.1)	453.5 (25.1)	0.701	< 0.001

[a] Continuous variables with normal distribution were reported as Mean(SD), continuous variables with abnormal distribution were reported as Median (IQR);

[b] r for Spearman correlation or Pearson correlation;

HR, Heart Rate

$$Sensitivity = \frac{\mathrm{T}P}{\mathrm{T}P + FN} \tag{2}$$

$$Specificity = \frac{\mathrm{T}N}{\mathrm{T}N + FP} \tag{3}$$

Results

Characteristics of participants

A total of 99 pregnant women who met the inclusion criteria were involved in this study, all of whom had single pregnancies. The demographics of the participants are presented in Table 1. No significant differences existed among different pregnant stages, expect for the participants' weight.

Reliability of WenXinWuYang

The results of the two repeated measures showed that the correlation coefficients of heart rate, PR interval, and QT interval were above 0.70, showing good test-retest reliability. The correlation coefficients of QRS duration, QTc interval were above 0.60, indicating acceptable reliability. Similar results were found among different pregnant stages, except for the first trimester, where the correlation coefficients of PR interval and QRS duration were below 0.60. The retest reliability of WenXinWuYang for continuous variables is summarized in Table 2.

The diagnostic consistency rates of the two repeated measures were 93.9% for all pregnant women, 97.0% for the first trimester group, 93.4% for the second trimester group, and 90.0% for the third trimester group. These results indicate a well acceptability of test-retest reliability in the diagnostic outcomes.

Validity of WenXinWuYang

Correlation and Bland-Altman analyses showed that the following absolute mean differences between the WenXinWuYang and 12-lead ECG: -0.4 ± 3.1 bpm for heart rate ($r = 0.957$); 14.6 ± 12.4 ms for PR interval ($r = 0.537$), 7.0 ± 8.9 ms for QRS duration ($r = 0.136$), 2.1 ± 12.0 for QT interval ($r = 0.774$), and 6.6 ± 16.5 for QTc interval ($r = 0.663$). Similar results were observed at different stages of pregnancy (Table 3, Table S1 in supplement). A strong correlation existed for heart rate and QT interval, moderate correlation for PR interval and QTc interval, and low correlation for QRS duration. Strong agreement existed in 87.9% of heart rate, 58.6% of PR interval, 88.9% of QRS duration, 88.9% of QT interval, and 79.8% of QTc interval. Moderate agreement existed in 10.1% of heart rate, 41.4% of PR interval, 11.1% of QRS duration, 11.1% of QT interval, and 16.2% of QTc interval, similar results were seen at different stages of pregnancy (Table S2 in Supplement). Bland-Altman plots for continuous variables among all pregnant women and those in different stages are shown in Fig. 1 and Figure S2-S4.

The diagnostic consistency rate of cardio rhythm between WenXinWuYang and the 12-lead ECG was 94.9%, showing a high level of agreement. The sensitivity was 84.2%, and the specificity was 97.5%. Table 4 describes the diagnostic performances of WenXinWuYang and 12-leads ECG (for the confusion matrix of diagnostic results from both devices, please see Table S3 in the Supplement). A total of 16 patients were diagnosed with abnormalities by both devices: 9 with Sinus Tachycardia (SNT), 4 with Sinus Arrhythmia (SNA), 1 with Right Bundle Branch Block (RBBB, ECG waveforms shown in Figure S5-A in the Supplement), 1 with Premature Ventricular Contraction (PVC, ECG waveforms shown in Figure S5-B in the Supplement), and 1 with PVC combined with SNA. There were 5 cases of inconsistent diagnostic results: 2 cases where the 12-lead ECG diagnosed SNA but WenXinWuYang recognized as normal, 2 cases where the 12-lead ECG diagnosed as normal but WenXinWuYang recognized as SNA, and 1 case where the 12-lead ECG diagnosed PVC combined with SNA but WenXinWuYang recognized as normal (ECG waveforms shown in Figure S6).

Discussion

To our knowledge, this is the first study to investigate the reliability and validity of the portable single-lead ECG device, including WenXinWuYang, among pregnant women in clinical practice. The correlation coefficients from the two repeated measures of heart rates and ECG

Table 3 The continuous variables validity of WenXinWuYang compared with 12-lead ECG[a]

	Parameters	12-lesd ECG	WenXinWuYang	r^b	*p*
Total	HR	80.0 (75.0, 91.5)	81.0 (75.0, 91.0)	0.957	< 0.001
	PR interval	134.0 (126.5, 149.5)	152.0 (145.0, 160.0)	0.537	< 0.001
	QRS duration	88.0 (84.0, 93.0)	96.0 (89.5, 98.5)	0.136	0.178
	QT interval	363.0 (360.0, 382.5)	374.0 (360.5, 381.0)	0.774	< 0.001
	QTc interval	435.0 (422.5, 440.0)	438.3 (22.4)	0.663	< 0.001
First trimester	HR	83.5 (78.0, 91.0)	82.5 (78.0, 90.5)	0.975	< 0.001
	PR interval	137.5 (129.0, 151.8)	152.0 (145.0, 160.0)	0.522	0.002
	QRS duration	90.0 (86.3, 96.0)	96.0 (89.0, 100.3)	0.345	0.046
	QT interval	360.0 (360.0, 372.8)	373.5 (361.3, 379.8)	0.542	< 0.001
	QTc interval	433.0 (423.8, 440.0)	438.5 (23.0)	0.759	< 0.001
Second trimester	HR	78.5 (73.8, 87.5)	78.0 (72.0, 85.3)	0.929	< 0.001
	PR interval	134.5 (123.5, 145.0)	152.5 (146.8, 160.3)	0.707	< 0.001
	QRS duration	88.5 (86.0, 91.0)	96.0 (92.0, 96.0)	0.031	0.865
	QT interval	374.0 (360.0, 388.3)	377.5 (358.8, 393.0)	0.853	< 0.001
	QTc interval	434.5 (418.8, 440.0)	435.6 (21.2)	0.658	< 0.001
Third trimester	HR	83.0 (76.0, 95.0)	82.0 (77.0, 95.0)	0.961	< 0.001
	PR interval	132.0 (128.0, 144.0)	147.0 (144.0, 158.0)	0.449	0.009
	QRS duration	86.0 (81.0, 91.0)	96.0 (90.0, 99.0)	0.074	0.684
	QT interval	363.0 (360.0, 376.0)	369.0 (362.0, 380.0)	0.823	< 0.001
	QTc interval	437.0 (424.0, 443.0)	440.6 (23.1)	0.586	< 0.001

[a] continuous variables with normal distribution were reported as Mean(SD), continuous variables with abnormal distribution were reported as Median (IQR);

[b] *r* for Spearman correlation;

HR, Heart Rate

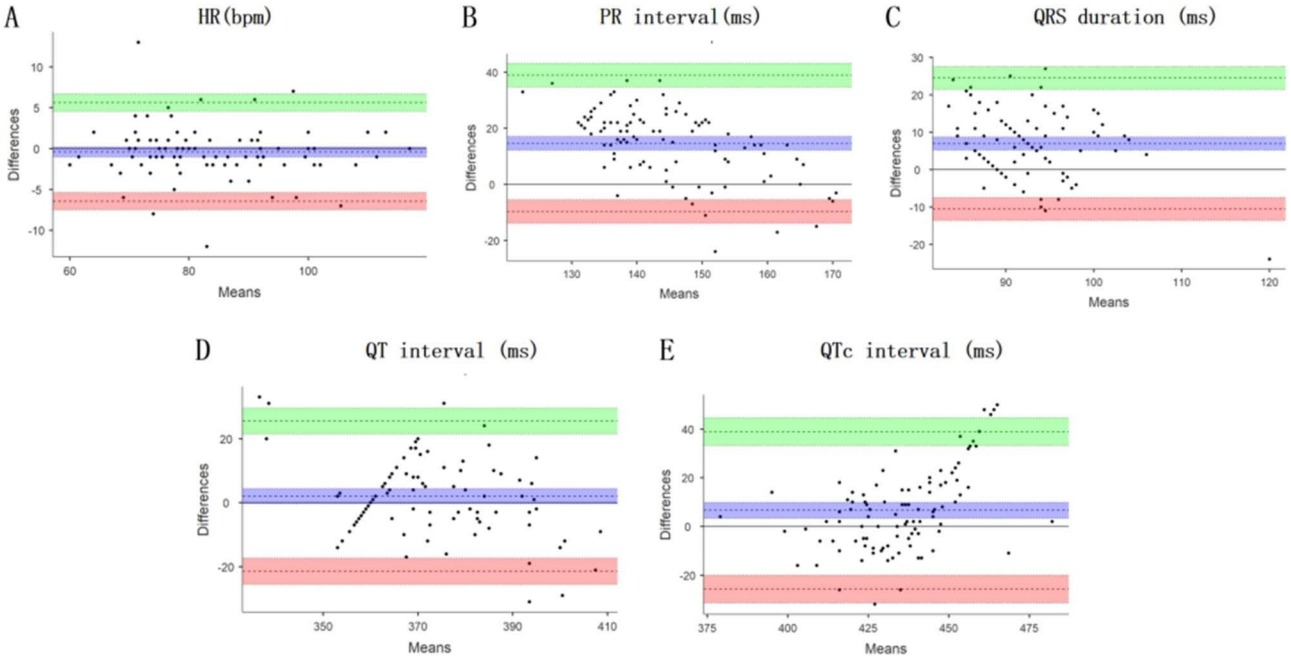

Fig. 1 Bland-Altman plots for measurement agreement between heart rate (**A**), PR interval (**B**), QRS duration (**C**), QT interval (**D**), and QTc interval (**E**) in WenXinWuYang and 12-lead ECG among all pregnant women ($n=99$). The middle dotted line represents the mean difference, the upper and lower dotted line represent + 1.96 and − 1.96 standard deviations, respectively

Table 4 The diagnostic performance of WenXinWuYang compared with 12-lead ECG(%)

	Accuracy	Sensitivity	Specificity
First trimester	94.1	83.3	96.4
Second trimester	93.8	71.4	100.0
Third trimester	97.0	100.0	96.2
Total	94.9	84.2	97.5

measurements, along with the consistency rate of diagnostic results, indicated that WenXinWuYang demonstrates good reliability in ECG measurements. Compared to the 12-lead ECG, WenXinWuYang showed moderate to high agreement on key parameters and high agreement in detecting arrhythmias. The results did not vary among different pregnant stages.

This study demonstrated the well reliability of the WenXinWuYang device in ECG measurement, although the correlation coefficients of QRS duration and QTc were slightly below 0.7. The reliability of detection is one of the key indicators to measure the quality of an equipment. However, few evidences were found to test the reliability of existing portable devices in clinical practice [5]. While the stability of the device during practical use may have been considered during its development [14], evaluating its reliability in real-world conditions remains an urgent need.

Although the final arrhythmia detection by WenXinWuYang relies on a deep learning algorithm rather than the basic ECG waveform intervals, we compared it with the 12-lead ECG to identify any potential advantages or disadvantages associated with the raw waveform tracings. This comparison helps determine whether these parameters can provide useful references for clinical practice.

Overall, our results show that WenXinWuYang can accurately measure heart rate and interval lengths in pregnant women without knowing their cardiac history. We evaluated the validity of the basic ECG intervals using correlation coefficients and agreement levels. Comprehensively, heart rate and QT interval collected by WenXinWuYang had high validity compared with the 12-lead ECG, and the results were similar or better than those of the single-lead Apple Watch (AW) ECG. Specifically, the mean difference in QT (2.1ms) was lower than the AW (-11.2ms), and the proportion of strong agreement (88.9%) was higher than AW (65.1%) [12]. However, it is important to note that the AW was evaluated in healthy people aged 18 years and older, not in pregnant women, and it was not tested simultaneously with the 12-lead ECG. This limits the comparability of the results. To improve comparability, more studies under similar conditions should be conducted in the future.

For QRS duration, there was strong agreement between the two devices, with results similar to those of the AW. However, the correlation coefficient (0.136) for WenXinWuYang was much lower than that of the AW (0.650) [12]. The scatterplot distribution of the Bland-Altman analysis helped explain why a lower correlation existed despite high agreement. This discrepancy may

be attributed to the limited sensitivity of the QRS duration algorithm, as 32.3% of the QRS duration values were concentrated at 89 ms and 35.4% at 96 ms. Given that the calculation of these values is within the normal range of deviations, we could still consider that the validity of QRS duration was good to some extent, but future work should be done to optimize the QRS duration algorithm.

The results of QTc interval in our study are also acceptable, with an absolute mean difference of 6.6 ± 16.5 ms and a strong agreement rate of 79.8%. This performance is better than that of the AW (-11.6 ± 27 ms) [12] and previous studies reported by Gropler (15.6 ± 12.7 ms) [30]. Moreover, the QTc interval performance was much better than that of the single-lead device mentioned by Charlotte, which was found to be inaccurate for measuring QTc interval [34]. Given that QTc interval prolongation may lead to potentially fatal cardiac arrhythmias [35], the validity of the QTc interval is of great significance.

In contrast, the results for the PR interval were somewhat weak, with a strong agreement rate of 58.6%, which was lower than that of the AW (83.3%) [12]. However, since the remaining 41.4% of cases demonstrated moderate agreement with the 12-lead ECG, the overall results were acceptable. Nonetheless, efforts should be made to improve the validity of the PR interval.

Our study found that the overall detection consistency between WenXinWuYang and the 12-lead ECG was 95.0%, comparable to that of the AW [12]. WenXinWuYang accurately identified all participants exhibiting sinus rhythm. Although its sensitivity for identifying normal ECGs was slightly lower (84.2%) compared to the Omron HeartScan (91.0%) as evaluated by Gerrit [36], most inconsistencies were related to the detection of sinus arrhythmia. Notably, participants with typical arrhythmias, such as PVC and RBBB, were almost all detected accurately. Several factors may account for these inconsistency. First, although there are established criteria for sinus arrhythmia, doctors may come to different conclusions based on their experience and interpretation of the ECG waveform. Second, evidence suggested that the validity of portable ECG devices can vary across different study populations [37]. Pregnancy is a special period with the increase of body load and change of hemodynamics in pregnant women, which may contribute to the increasing inconsistency of detection. Future deep learning models for ECG disease detection should pay more attention on pregnant women. Third, the small sample size and insufficient statistical efficiency in this study may have also influenced the results. Studies with larger sample sizes are needed to verify the findings.

Subgroup analyses of different pregnant stage showed similar results which indicated that WenXinWuYang could be used throughout pregnancy. While arrhythmias during pregnancy are not uncommon, most are benign and treatable. However, given the risks to both the mother and fetus, regular ECG screening is still necessary [38]. As a portable ECG device, the findings of this study provide evidence supporting reliability and detection validity of WenXinWuYang, demonstrating its potential in screening for arrhythmias in pregnant populations.

Strengths and limitations

The primary strength of this study lies in its comprehensively analysis of WenXinWuYang's reliability and validity. First, we conducted two repeated tests to verify the device's reliability. Second, we compared the result with the gold standard 12-lead ECG in clinical practice, and for the first time, we performed simultaneous measurements with WenXinWuYang and the 12-lead ECG to ensure the consistency of the results. Third, we not only compared heart rate and interval indicators with the 12-lead ECG but also evaluated diagnostic outcomes. Finally, we assessed the validity of portable ECG devices in pregnant population, taking into account the effects of different pregnant stages for the first time.

This study has some limitations. First, the simultaneous measurement of the two devices had slight signal interference, but this does not affect the ECG detection conclusions. On the contrary, it helped ensure greater consistency between the measurement results of the two devices. Second, this study was conducted among general pregnant women without knowing their arrhythmias. Due to the small sample size and the low incidence of arrhythmias within the study population, the statistic power of the device in identifying a specific category of arrhythmias may be inadequate. However, the objective of this study was not focus on the assessment of detection validity for a specific type of arrhythmia (such as atrial fibrillation). The high reliability and validity demonstrated by WenXinWuYang suggest that it holds promise for detecting various types of arrhythmia disease. Third, we compared the ECG intervals recorded by WenXinWuYang with the average or representative values obtained from all 12 leads of 12-lead ECG, rather than with the results from a single lead. While this may slightly reduce comparability, the use of comprehensive indices offers greater practical significance for clinical diagnosis, making this comparison approach feasible.

Conclusion

This comprehensive evaluation results supported the reliability and validity of WenXinWuYang in detecting arrhythmias among pregnant women, showing moderate to strong agreement in heart rate and ECG intervals, as well as high validity in diagnostic results. Further studies with large sample sizes and different population are needed to verify these findings.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-025-02952-6 .

Supplementary Material 1

Acknowledgements
The authors acknowledge all the pregnant women who took their time to participate in this study, Qinghao Zhao (Department of Cardiology, Peking University People's Hospital, Beijing, China) and Xinxin Di (Department of Electrocardiogram, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, China) for their comments on the results of the study.

Author contributions
HW (the first author), HW (the corresponding author), and SH designed the research, HW (the first author), JW, WJ, and SD performed data collection. DZ and SG conducted data curation. HW and JW analyzed and processed the data, wrote the manuscript. All authors reviewed the results and approved the final version of the manuscript.

Funding
This work is funded by the research fund of the Peking University Health Science Center-Weifang Joint Research Center for Maternal and Child Health (PKUWF-T04), National Natural Science Foundation of China (No.62102008) and Beijing Natural Science Foundations (QY23040). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Data availability
Research data are not publicly available but can be obtained from the corresponding author on request.

Declarations

Ethics approval and consent to participate
All procedures and methods were performed in accordance with the relevant guidelines and regulations. Ethics approval of this study was granted by the Medical Ethics Committee of Weifang Maternal and Child Health Hospital (Number 2023 – 211) prior to the beginning of this study. All participants provided informed consent for data collection and analyses. All data were analyzed anonymized.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]National Institute of Health Data Science, Peking University, Peking University Health Science Center- Weifang Joint Research Center for Maternal and Child Health, Beijing, China
[2]Weifang Maternal and Child Health Hospital, Peking University Health Science Center-Weifang Joint Research Center for Maternal and Child Health, Weifang, Shandong Province, China
[3]HeartVoice Medical Technology, Hefei, China
[4]Department of Maternal and Child Health, School of Public Health, Peking University, Peking University Health Science Center-Weifang Joint Research Center for Maternal and Child Health, Beijing, China
[5]National Institute of Health Data Science, Peking University, Beijing, China
[6]Institute of Medical Technology, Peking University Health Science Center, Beijing, China
[7]Department of Emergency Medicine, Peking University First Hospital, Beijing, China

Received: 19 May 2024 / Accepted: 25 February 2025
Published online: 03 March 2025

References

1. Deng K, Collaborators G. C.O.D: global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the global burden of disease study 2021. Lancet. 2024;403(0440):2100–2132.
2. Mariachiara DC, Honor B, Thomas G, Lisa H, Chodziwadziwa K, Diana VM, et al. World health report 2023: confronting the world's number one killer. Geneva, Switzerland: World Heart Federation; 2023.
3. WHO. About cardiovascular diseases. Who.int. 2017 [cited 2017 Mar 21]. Available from http://www.who.int/cardiovascular_diseases/about_cvd/en/
4. Spatz ES, Ginsburg GS, Rumsfeld JS, Turakhia MP. Wearable digital health technologies for monitoring in cardiovascular medicine. N Engl J Med. 2024;390(4):346–56.
5. Bansal A, Joshi R. Portable out-of-hospital electrocardiography: a review of current technologies. J Arrhythm. 2018;34(2):129–38.
6. Mehta DD, Nazir NT, Trohman RG, Volgman AS. Single-lead portable ECG devices: perceptions and clinical accuracy compared to conventional cardiac monitoring. J Electrocardiol. 2015;48(4):710–6.
7. Haberman ZC, Jahn RT, Bose R, Tun H, Shinbane JS, Doshi RN, et al. Wireless smartphone ECG enables large-scale screening in diverse populations. J Cardiovasc Electrophysiol. 2015;26(5):520–6.
8. Lau JK, Lowres N, Neubeck L, Brieger DB, Sy RW, Galloway CD, et al. iPhone ECG application for community screening to detect silent atrial fibrillation: a novel technology to prevent stroke. Int J Cardiol. 2013;165(1):193–4.
9. Torfs T, Smeets CJ, Geng D, Berset T, Van der Auwera J, Vandervoort P, et al. Clinical validation of a low-power and wearable ECG patch for long term full-disclosure monitoring. J Electrocardiol. 2014;47(6):881–9.
10. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, et al. Passive detection of atrial fibrillation using a commercially available smartwatch. Jama Cardiol. 2018;3(5):409–16.
11. Dobbs WC, Fedewa MV, MacDonald HV, Holmes CJ, Cicone ZS, Plews DJ, et al. The accuracy of acquiring heart rate variability from portable devices: a systematic review and meta-analysis. Sports Med. 2019;49(3):417–35.
12. Saghir N, Aggarwal A, Soneji N, Valencia V, Rodgers G, Kurian T. A comparison of manual electrocardiographic interval and waveform analysis in lead 1 of 12-lead ECG and Apple watch ECG: a validation study. Cardiovasc Digit Health J. 2020;1(1):30–6.
13. Garabelli P, Stavrakis S, Albert M, Koomson E, Parwani P, Chohan J, et al. Comparison of QT interval readings in normal sinus rhythm between a smartphone heart monitor and a 12-lead ECG for healthy volunteers and inpatients receiving sotalol or dofetilide. J Cardiovasc Electrophys. 2016;27(7):827–32.
14. Zompanti A, Sabatini A, Grasso S, Pennazza G, Ferri G, Barile G, et al. Development and test of a portable ECG device with dry capacitive electrodes and driven right leg circuit. Sensors. 2021;21(8):2777.
15. Ding EY, Marcus GM, McManus DD. Emerging technologies for identifying atrial fibrillation. Circul Res. 2020;127(1):128–42.
16. Kamga P, Mostafa R, Zafar S. The use of wearable ECG devices in the clinical setting: a review. Curr Emerg Hosp Med Rep. 2022;10(3):67–72.
17. Nie GK, Zhu JB, Tang GZ, Zhang DY, Geng SJ, Zhao QH, Hong SD. A [preprint] eview of deep learning methods for [preprint]hotoplethysmography data [Preprint], arXiv. 2024.
18. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Injury. 2011;42(3):236–40.
19. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. Am J Health-System Pharm. 2008;65(23):2276–84.
20. Bannigan K, Watson R. Reliability and validity in a nutshell. J Clin Nurs. 2009;18(23):3237–43.
21. HU S-S. Report on cardiovascular health and diseases in China 2021:an updated summary. J Geriatric Cardiol. 2023;20(6):399–430.
22. Ma LY, Wang ZW, Fan J, Hu SS. Key points of China cardiovascular health and disease report 2021. J Chin Gen Pract. 2022;25(27):3331–46.
23. Hong SD, Fu ZJ, Zhou RB, Yu J, Li YK, Wang K, et al. CardioLearn: a cloud deep learning service for cardiac disease detection from electrocardiogram. In Companion Proceedings of the Web Conference 2020 (WWW'20). Association for Computing Machinery, New York, NY, USA, 148–152.
24. Fu Z, Hong S, Zhang R, Du S. Artificial-intelligence-enhanced mobile system for cardiovascular health management. Sensors. 2021;21(3):773.
25. HeartVoice (2024, August 10). Caring for human heart health with AI. https://www.heartvoice.com.cn/en/about.html
26. O'Kelly AC, Michos ED, Shufelt CL, Vermunt JV, Minissian MB, Quesada O, et al. Pregnancy and reproductive risk factors for cardiovascular disease in women. Circul Res. 2022;130(4):652–72.

27. Limongelli G, Rubino M, Esposito A, Russo M, Pacileo G. The challenge of cardiomyopathies and heart failure in pregnancy. Curr Opin Obst Gynecol. 2018;30(6):378–84.
28. Bett GCL. Hormones and sex differences: changes in cardiac electrophysiology with pregnancy. Clin Sci. 2016;130(10):747–59.
29. Brener A, Briller J. Cardiovascular testing and imaging in pregnant women. Cardiol Clin. 2021;39(1):21–32.
30. Gropler MRF, Dalal AS, Van Hare GF, Silva JNA. Can smartphone wireless ECGs be used to accurately assess ECG intervals in pediatrics? A comparison of mobile health monitoring to standard 12-lead ECG. PLoS ONE. 2018;13(9):e0204403.
31. Nadimi-Shahraki MH, Varzaneh ZA, Zamani H, Mirjalili S. Binary starling murmuration optimizer algorithm to select effective features from medical data. Appl Sciences-Basel. 2023;13(1):564.
32. Xu JL, Mei XY, Chen YF, Wan XK. An effective premature ventricular contraction detection algorithm based on adaptive template matching and characteristic recognition. Signal Image Video Process. 2024;18(3):2811–8.
33. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol. 2003;56(11):1129–35.
34. Bekker CL, Noordergraaf F, Teerenstra S, Pop G, van den Bemt BJF. Diagnostic accuracy of a single-lead portable ECG device for measuring QTc prolongation. Ann Noninvasive Electrocardiol. 2020;25(1):e12683.
35. Al-Khatib SM, LaPointe NM, Kramer JM, Califf RM. What clinicians should know about the QT interval. JAMA. 2003;289(16):2120–7.
36. Kaleschke G, Hoffmann B, Drewitz I, Steinbeck G, Naebauer M, Goette A, et al. Prospective, multicentre validation of a simple, patient-operated electrocardiographic system for the detection of arrhythmias and electrocardiographic changes. Europace. 2009;11(10):1362–8.
37. Giebel GD, Gissel C. Accuracy of mHealth devices for atrial fibrillation screening: systematic review. Jmir Mhealth Uhealth. 2019;7(6):e13641.
38. Li J-M, Nguyen C, Joglar JA, Hamdan MH, Page RL. Frequency and outcome of arrhythmias complicating admission during pregnancy: experience from a high-volume and ethnically-diverse obstetric service. Clin Cardiol. 2008;31(11):538–41.

Publisher's note