Proceedings of the 2nd World Congress on Industrial Engineering and Operations Management
Windsor, Canada, October 14-16, 2025

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/WC02.20250036
Published: October 14, 2025

Evaluating the Statistical Reliability of Wearable Cardiovascular Health Monitoring Devices Using Gage R and R

Zan Chou
High School Student
Huaxing High School
Taipei, Taiwan
s11231062@hhhs.tp.edu.tw, zion97chou2008@gmail.com

Ni-Hsi Yeh
High School Student
Taipei City Fanghe Experimental High School
11230021@fhehs.tp.edu.tw, nissi109033@gmail.com

Pin-Jen Lai
High School Student
Kang Chiao International School Linkou Campus
Taipei, Taiwan
LS12195@stu.kcislk.ntpc.edu.tw, nicolelai0511@gmail.com

Abstract

As the global population continues to age, cardiovascular diseases (CVDs) are becoming increasingly prevalent. In Taiwan, the country entered an aging society as early as 2003 and is projected to become a super-aged society by 2025. Effective management of CVDs is essential to reducing the burden on the healthcare system. Due to their accessibility and practicality, wearable health devices have emerged as useful tools for daily cardiovascular monitoring. However, concerns remain regarding their measurement stability and accuracy. This study applies Measurement System Analysis (MSA) to evaluate the stability of wearable cardiovascular health monitoring devices. Using Gage Repeatability and Reproducibility (Gage R&R), misclassification probability analysis, and two-way ANOVA with interaction effects, we identify potential sources of measurement error. Considering individual health variability, we aim to identify the most suitable user groups in order to reduce delayed treatment and medical resource waste. From a human factors perspective, we also provide recommendations for consumers when selecting appropriate devices. Additionally, the study proposes design and calibration improvements to help manufacturers enhance device stability. By understanding measurement variability and its contributing factors, this research aims to improve the accuracy of wearable cardiovascular health monitoring. Ultimately, the findings contribute to better consumer guidance, more reliable product development, and reduced healthcare strain in super-aged societies.

Keywords
Cardiovascular health, wearable monitoring devices, Gage Repeatability and Reproducibility, STEAMS.

1. Introduction

Taiwan is rapidly transitioning into a super-aged society, with the population aged 65 and above projected to exceed 20% by 2025 (National Development Council, 2022). Cardiovascular and hypertensive diseases ranked as the top two causes of death in Taiwan in 2024 (Ministry of Health and Welfare, 2024), highlighting the urgent need for effective daily health management. As mortality from cardiovascular diseases increases, public attention to health monitoring has gradually risen. The widespread adoption of smart health-monitoring watches has further accelerated this trend, creating higher expectations for the accuracy and reliability of health measurement results. Reliable and accessible health monitoring has thus become increasingly important in modern society.

Some conditions—such as paroxysmal atrial fibrillation, white coat or masked hypertension, morning hypertension, and blood pressure variability—cannot be captured by single, non-reproducible measurements. Long-term monitoring better reflects real-life physiological states, enabling earlier risk detection, complication prevention, and timely intervention. Although professional medical instruments provide highly precise measurements, their prohibitive costs make them inaccessible to the general public. Consequently, this study focuses on wearable devices, particularly smartwatches, which, due to their portability, functional integration, and real-time monitoring capabilities, have emerged as essential tools for personal health management (Wang et al., 2022). Nonetheless, mid- to low-priced wearable devices currently exhibit inconsistent stability and accuracy when measuring physiological parameters such as heart rate and blood pressure, raising concerns about their reliability. To address this, the present study emphasizes stability as a primary research focus, proposing that only stable measurements allow accuracy to be further enhanced through coefficient adjustments or calibration. In collaboration with (manufacturer name), we tested the Guider ECG SpO2 Smart Watch 800Z, examining both device-related and human factors that may influence measurement outcomes. Furthermore, this study proposes a methodology for evaluating the measurement accuracy of wearable devices, aiming to facilitate the popularization of smart health devices and provide concrete guidance for future technological developments in the health monitoring field.

2. Literature Review

Clark et al. (2016) reported that a persistent inter-arm blood pressure difference exceeding 10 mmHg is a marker of vascular pathology and increased cardiovascular risk. As the global population rapidly advances toward a super-aged society, the incidence of cardiovascular disease rises concomitantly with age. Studies have demonstrated that for every 10 beats per minute increase in resting heart rate (RHR), there is a significant elevation in both all-cause and cardiovascular mortality risk (Zhang et al., 2016; Woodward et al., 2014). Similarly, a 10 mmHg increase in systolic blood pressure (SBP) is associated with approximately a 17% increase in coronary heart disease mortality (van den Hoogen et al., 2000). These findings emphasize the critical importance of daily monitoring of heart rate and blood pressure in managing the health of older adults.

Wearable devices capable of measuring heart rate and blood pressure have thus emerged as convenient tools for the elderly, potentially playing a vital role in early disease detection and risk management. However, the utility of such devices hinges on their measurement stability—the consistency of data output across repeated measurements within a short timeframe. Even if absolute accuracy compared to clinical-grade instruments is lacking, minimal variation between repeated measures allows for meaningful health indicators, which can be further refined through calibration or parameter adjustments. Furthermore, hypertension arises from multifactorial causes including vascular stiffening, autonomic nervous system dysregulation, and genetic predisposition (Franklin et al., 1997; Grassi et al., 2015). While physiological hypotension is common among healthy individuals, orthostatic hypotension—characterized by a significant drop in blood pressure upon standing—may indicate autonomic dysfunction (Freeman et al., 2011). Therefore, consistent and stable monitoring of heart rate and blood pressure trends is especially meaningful for early identification of abnormal patterns in the elderly population.

2.1 Photoplethysmography (PPG) in Wearable Cardiovascular Healthcare (CVH) Devices

According to Li et al., "PPG (Photoplethysmography) is a non-invasive technology that uses a light source and a photodetector at the skin's surface to measure the volumetric variations in blood circulation" (2023). Measurement is typically taken over the arterial regions, of which rhythmic pulsations— reflecting the user's cardiac cycle— result in periodic changes in light absorption (Kim and Baek 2023). Popularly adopted in CVH wearable devices, detected trends in pulse intervals have been used to monitor a number of biometric variables, such as heart rate, heart rate variability, blood pressure, and blood oxygen saturation, etc. However, it is also notably hypersensitive to external influence, such as motion, humidity, ambient light, skin color, etc. (Li et al. 2013). To better understand the fundamental issues behind PPG's strengths and weaknesses, it is necessary to examine the composition and working principles of its technology.

2.2 Composition and Working Principles

Among commercial wearable CVH devices, wrist-worn smart devices such as the Apple Watch Series 4+ or the Fitbit Series have emerged with the most prevalent adoption, likely due to their multifunctional design beyond healthcare purposes (Kim and Baek 2023). In this context, the integration of PPG technology relies on the measurement of wrist artery blood flow, of which purpose is reflected by sensor components and layout on the apparatus (Figure 1).

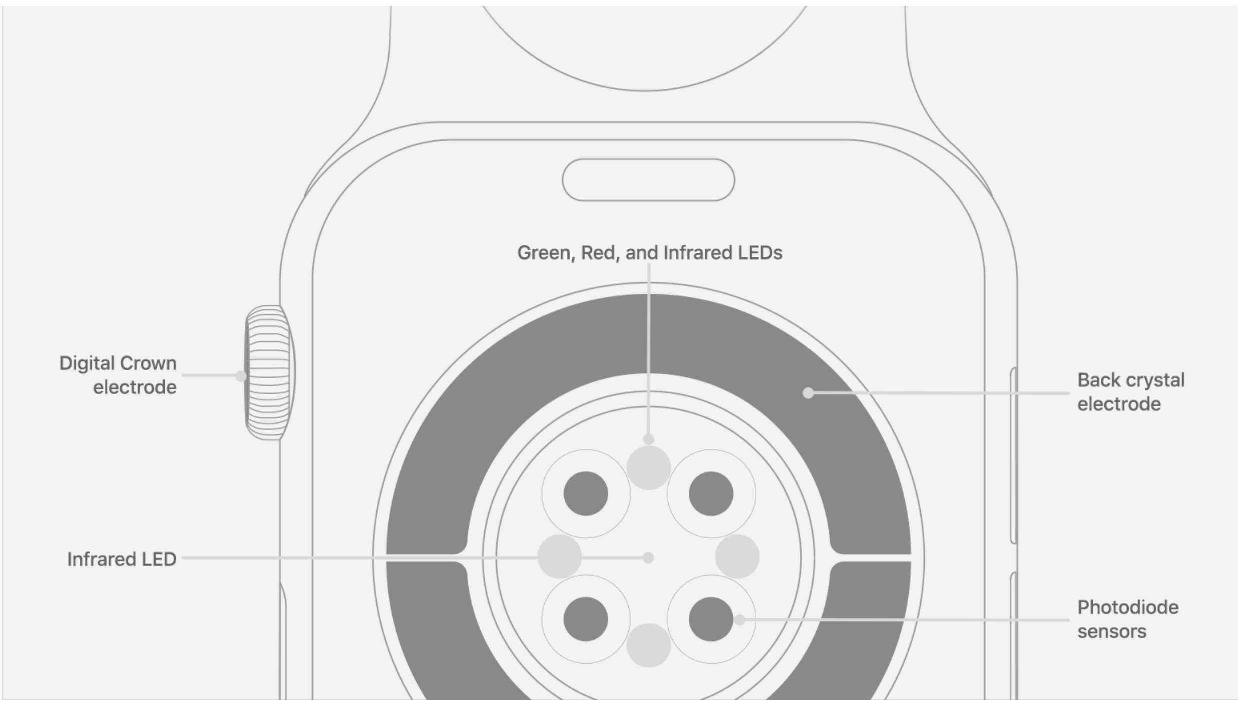

Figure 1. Apple Watch Series 6 sensor layout (Li et al. 2023). Photodiode sensors are arranged between green, red, and infrared LEDs to maximize light absorption. Contact with the digital crown electrode, connected to the back crystal electrode, allows the watch to send an electrical signal and record the rhythm of the heart.

2.2.1 PPG Sensor Components

In wrist-worn CVH devices, the mainstream composition for PPG sensing is a combination of LED lights paired with one or more photodiode sensors (PD). With the rise of commercial healthcare wearables, this combination came in the form of LED-PD compounds with analog front ends (AFE) (Kim and Baek 2023). Specifically optimized for health wearable use, compact size and energy efficiency of the integrated module allowed smoother incorporation of PPG technology into smaller-sized wearables, whereas the AFE circuit was crucial in improving assessment precision by preprocessing LED-PD signals to eliminate external interference. While adopted by most commercially available CVH wearables for its convenience and reliability, its greatest limitation lies in its inflexibility during device design (Kim and Baek 2023).

The variety of PPG layout designs for wrist-worn CVH devices reflects the need for varying wavelength combinations to target different physiological metrics. A key purpose for layout design is to minimize crosstalk noise while optimizing measurement signals, therefore requiring flexibility in component orientation to achieve versatile combinations. However, the popularization of standardized LED-PD compounds constraints this flexibility, limiting customization potential. Therefore, to best meet detection requirements, layout designers strategically orient different wavelength LEDs to enhance light-tissue interaction (Kim and Baek 2023).

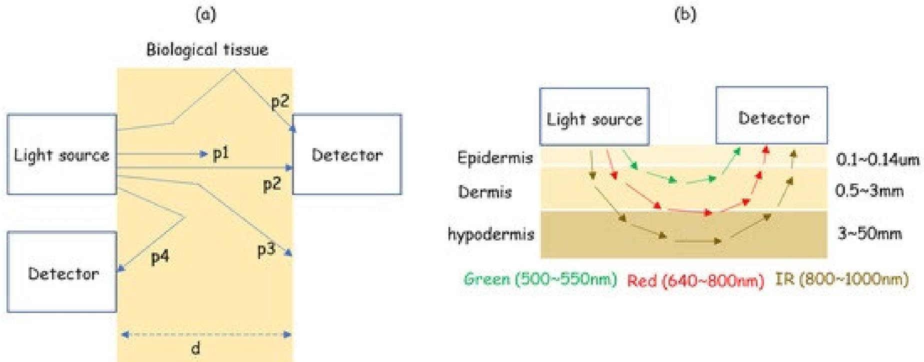

Figure 2. Measurement model of reflectance type PPG through the skin microvascular layer, emphasis on different wavelengths (Kim and Baek 2023).

It is important to recognize the optical properties of specific wavelengths when interacting with various biological and environmental factors during measurement (Figure 2). As seen in Figure 2, the majority of CVH wearables use a combination of green, red, and infrared wavelengths to monitor blood and tissue characteristics. Green LED lights are noted for their high absorptivity by hemoglobin, which allows them enhanced detection of changes in blood flow, as well as their high signal-to-noise ratio in conducting motion-prone measurements, making them suitable for wearables (see Figure 3(b)) (Lee et al. 2013; Tamura 2019). However, their susceptibility to melanin introduces noise across a spectrum of skin tones. While this does not bother their ability to measure relatively shallow wrist arteries, they struggle to detect deeper veins and capillaries needed for blood oxygen measurements (Lee and Hahm, 2016).

Infrared light, on the other hand, is capable of deeper penetration into vascular tissue and is less affected by skin tones or ambient lights. It is vulnerable to motion-related noise factors and offers less signal amplitude than green light, making the two highly complementary. Red light purposes as an intermediary of the two, balancing external noise for measurement accuracy (Lee and Hahm 2016). As a result, device manufacturers may selectively emphasize certain wavelengths over others depending on their intended application: for instance, green light has been known to excel in pulse detection, and IR for oximetry (Lee et al. 2017; Sun and Thakor 2015). Effective wavelength selection paired with clever orientation enhances device precision by balancing signal penetration depth, sensitivity to external influencers to attain clearer, stronger CVH signals.

2.2.2 Component Layout Influence on Measurement Reliability

Acknowledged for their clinical importance and technological feasibility, commercial CVH devices often feature heart rate and blood pressure measuring functions. This focus is particularly relevant pertaining to elderly demographics, whose weakened organs could contribute to issues such as heart failure, arrhythmia, and hypertension.

Given the elderly population's vulnerability to cardiovascular health-related issues, it is crucial to minimize assessment errors to accurately guide what could be early diagnosis. In order to optimize CVH devices to be more acutely responsive to these signals, manufacturers have to recognize how age-

related factors may impact data quality. Physiological changes, such as vascular and skin aging, could result in lower capillary perfusion or darker skin tones that would directly influence light absorption from the PPG sensor.

Heart rate, rhythm, and blood pressure can all be identified by waveforms for blood absorbance in the arteries. When the heart relaxes, known as systole, arterial pressure drops; conversely, diastole is when heart muscles contract to send blood through vessels, allowing blood pressure to peak. As the heart circulates through a complete systole-diastole cycle, commonly put as a 'pulse', it creates a periodic change in blood absorbance (Allen et al. 2020, Charlton et al., 2021, . PPG sensors estimate the consistency and number of intervals between diastolic peaks to determine heart rhythm and rate, while the risetime and amplitude of said peaks are then used to determine blood pressure. However, relying on a single physiological parameter to assess CVH, without accounting for how age-related changes may obscure measurements, could pose significant risk to result accuracy.

Normal vascular aging results in stiffened arteries and thinner skin, both factors which have been determined to create noise in PPG-based CVH measurements. Hardened arteries could block blood from freely flowing through the vessels, contributing to ambiguous peaks on the waveform and higher arterial pressure. On the other hand, aged skin layers lose their uniform collagen and epidermal networks, causing light to scatter instead of appropriately reflecting to the surface. As a solution, Charlton et al. have proposed using PPG to determine vascular age, to which Allen et al.'s findings support by citing an inverse relationship between pulse risetime and age. Yet even now, the majority of commercially available CVH monitoring devices lack the specificity to tailor to elderly consumers' needs.

3. Methodology

In this study, JMP version 18 software was used to perform data analysis and conduct ANOVA calculations.This study applies a Measurement System Analysis (MSA) approach, utilizing a two-way ANOVA with interaction to examine the impact of two factors and their interaction on measurement accuracy.

The two factors are:

(1) Patient samples — To evaluate whether the device can accurately record each subject's systolic pressure, diastolic pressure, and heart rate, ensuring precision in identifying and measuring physiological values. Based on established clinical thresholds, patients are categorized as "conforming" (with normal clinical values) or "non-conforming" (with abnormal values) to assess the device's performance across both homogeneous and heterogeneous groups.

(2) Left vs. right hand — to assess whether the measurement results are affected by human-related variability. According to a certain medical theory, significant differences between left and right hand readings may indicate endocrine imbalance risk in the subject.

The interaction effect, while not the primary focus, plays a necessary role in this analysis. Since each subject's physiological baseline differs, ignoring the interaction may result in those effects being incorrectly attributed to repeatability (machine error), leading to inaccurate interpretations of the measurement system's performance.

3.1 Variability / Attribute Gauge

This study uses the P/T (Precision-to-Tolerance) ratio to evaluate the device's measurement variability relative to the clinical tolerance limits for hypertension, hypotension, and heart rate, which are defined as 90–140 mmHg, 60–90 mmHg, and 60–100 bpm, respectively. The P/T ratio directly reflects the proportion of the allowed tolerance range consumed by the measurement system's variability, making it a suitable metric to determine whether the device can reliably distinguish between conforming and non-conforming readings. Unlike the P/TV ratio, which incorporates part-to-part variation, the P/T ratio focuses solely on measurement precision, aligning more closely with the objectives of this study. Accordingly, part-to-part variation is not addressed in this analysis.

In addition, Gauge R&R and attribute gauge analyses were conducted to assess measurement repeatability and classification accuracy, providing a comprehensive evaluation of the device's performance. These analyses allow for the identification of variation attributable to repeatability error, total gauge error, and part-to-part variation, thereby offering valuable insights into potential sources of measurement error and the instrument's sensitivity in distinguishing between different sample parts.

Through an analysis of quality and process variability, as well as attribute gauge analysis, a variability chart and a Gauge Repeatability and Reproducibility (Gauge R&R) assessment were obtained. These tools help identify the proportion of variation attributable to repeatability error, total gauge error (Gauge R&R), and part-to-part variation. Such data provide valuable insights into the potential sources of measurement error and the instrument's sensitivity in distinguishing between different sample parts.

According to the Measurement Systems Analysis Reference Manual (3rd Edition, 2002) published by the Automotive Industry Action Group (AIAG) and based on the Six Sigma methodology, when the %Gauge R&R or any single source of variation exceeds the acceptable threshold of 30%, it indicates a high risk that more than 30% of the data may be misclassified, thereby significantly compromising the accuracy of data analysis.In such cases, instrument calibration and targeted analysis of the error sources are required to identify and correct the root causes.

Repeatability refers to the measurement error caused by the instrument itself when repeatedly measuring the same sample under identical conditions. A higher percentage of repeatability indicates poorer instrument stability and serves as an important indicator for assessing the reliability of a measurement system. Total gauge error (Gauge R&R) is the square root of the sum of the squares of repeatability and reproducibility, used to evaluate the overall error of the measurement system; a higher Gauge R&R signifies greater instability of the system. On the other hand, part-to-part variation reflects the true differences between different samples. Unlike measurement errors, a high part-to-part variation indicates that the samples are clearly differentiated and representative. Additionally, the higher the proportion of part-to-part variation, the smaller the relative impact of measurement error on overall variation, thereby reducing its interference with the analysis results. It should be noted that if part-to-part variation decreases in future studies, a high repeatability component may significantly affect the ability to distinguish true sample differences from measurement noise. Therefore, analyzing the sources of high repeatability is crucial for improving instrument stability and measurement accuracy.

3.2 Misclassification Probabilities

Misclassification probabilities are used to evaluate the likelihood that a measurement system incorrectly classifies samples as either acceptable or defective. In this study, the analysis focuses on misclassification related to systolic pressure, diastolic pressure, and heart rate values falling inside or outside the defined acceptability categories, with probabilities of alpha error, where the good part is falsely rejected, and beta error, where the bad part is falsely accepted.

In this study, assessing the potential for misclassification is of critical importance, as inaccurate measurements may lead to delays in medical intervention and the inappropriate allocation of healthcare resources. Regarding alpha error, misclassifying a normal value as abnormal may lead to unnecessary consumption of medical resources and undermine the credibility of subsequent abnormal readings, causing both users and healthcare professionals to question their validity. More critically, regarding beta error, misclassifying an abnormal value as normal may delay diagnosis and treatment, placing the individual at significant health risk and, in severe cases, even endangering their life.

A threshold of 20% is used in this study as the critical limit for misclassification. If any of the above probabilities exceed 20%, it indicates poor repeatability of the system, and such results should be flagged as unreliable and subject to correction or calibration (Table 1).

Table 1. Misclassification Probabilities Based on Tolerance Limits and Grand Mean

Misclassification Probabilities

Lower Tolerance = 60, Upper Tolerance = 100, Grand Mean = 71.96667

Description	Probability
P(Good part is falsely rejected)	0.03890073
P(Bad part is falsely accepted)	0.15295962
P(Part is good and is rejected)	0.03334087
P(Part is bad and is accepted)	0.02186166
P(Part is good)	0.85707561

4. Analysis

4.1 The Analysis of heart rate

According to Table 1, the repeatability component accounted for 44.22% of the total variation, representing the largest contributor to the overall measurement variance. Rather than suggesting instrument malfunction, this result highlights the capability of the applied methodology to identify stability-related factors within the measurement process. The corresponding total gauge error (Gauge R&R) was 47.72% (Table 2), indicating that the proposed analytical approach effectively captures the degree of variability inherent in repeated measurements.Meanwhile, the part-to-part variation accounted for 164.52%, demonstrating that the tested samples exhibited substantial and representative differences. This further validates the suitability of the analytical framework for distinguishing true sample variability from instrument-related error. In this context, the high part-to-part variation reduces the relative influence of repeatability error, thereby minimizing its interference with overall analysis outcomes.It is worth noting that if part-to-part variation were to decrease in future datasets, a higher repeatability ratio could potentially obscure the differentiation between genuine sample variation and measurement noise. Hence, the methodological framework adopted in this study provides not only a diagnostic assessment of current measurement reliability but also a structured basis for future stability improvement and instrument evaluation.

Table 2. Gauge R&R Analysis of Measurement Variation and Its Components

Gauge R&R

Measurement Source		Variation (6*StdDev)	% of Tolerance		which is 6*sqrt of
Repeatability	(EV)	17.686605	44.22	Equipment Variation	V(Within)
Reproducibility	(AV)	7.177737	17.94	Appraiser Variation	V(arm) + V(arm*Run Order病患 2)
arm		2.574104	6.44		V(arm)
arm*Run Order病患 2		6.700291	16.75		V(arm*Run Order病患 2)
Gauge R&R	(RR)	19.087585	47.72	Measurement Variation	V(Within) + V(arm) + V(arm*Run Order病患 2)
Part Variation	(PV)	65.808987	164.52	Part Variation	V(Run Order病患 2)
Total Variation	(TV)	68.521228	171.30	Total Variation	V(Within) + V(arm) + V(arm*Run Order病患 2) + V(Run Order病患 2)

Table 3. Variability of 17 Participants's heart rates tested via Guider ECG SpO_2 Smart Watch 800Z.

Variability Gauge Analysis for 心跳

Gauge R&R

Measurement Source		Variation (6*StdDev)	% of Tolerance		which is 6*sqrt of
Repeatability	(EV)	18.156696	45.39	Equipment Variation	V(Within)
Reproducibility	(AV)	7.158867	17.90	Appraiser Variation	V(arm) + V(arm*病患)
arm		2.480024	6.20		V(arm)
arm*病患		6.715569	16.79		V(arm*病患)
Gauge R&R	(RR)	19.517044	48.79	Measurement Variation	V(Within) + V(arm) + V(arm*病患)
Part Variation	(PV)	62.320150	155.80	Part Variation	V(病患)
Total Variation	(TV)	65.304794	163.26	Total Variation	V(Within) + V(arm) + V(arm*病患) + V(病患)

From this Table 3, we can observe that the repeatability error is 45.39% > 30%, indicating that the error caused by the instrument itself is relatively high, representing the largest factor affecting instrument stability.Looking at the Arm factor, the reproducibility between the left and right hands is about 6%, which demonstrates that in this analysis, the influence of human-related factors such as endocrine imbalance is minimal. For heart rate measurements, the main source of error stems from the stability of the device rather than human variability.This corresponds to the Figure 3 below, where the trend of left- and right-hand samples (blue trend line) closely aligns with the overall mean level (purple line), suggesting that human factors have little impact on the data analysis.

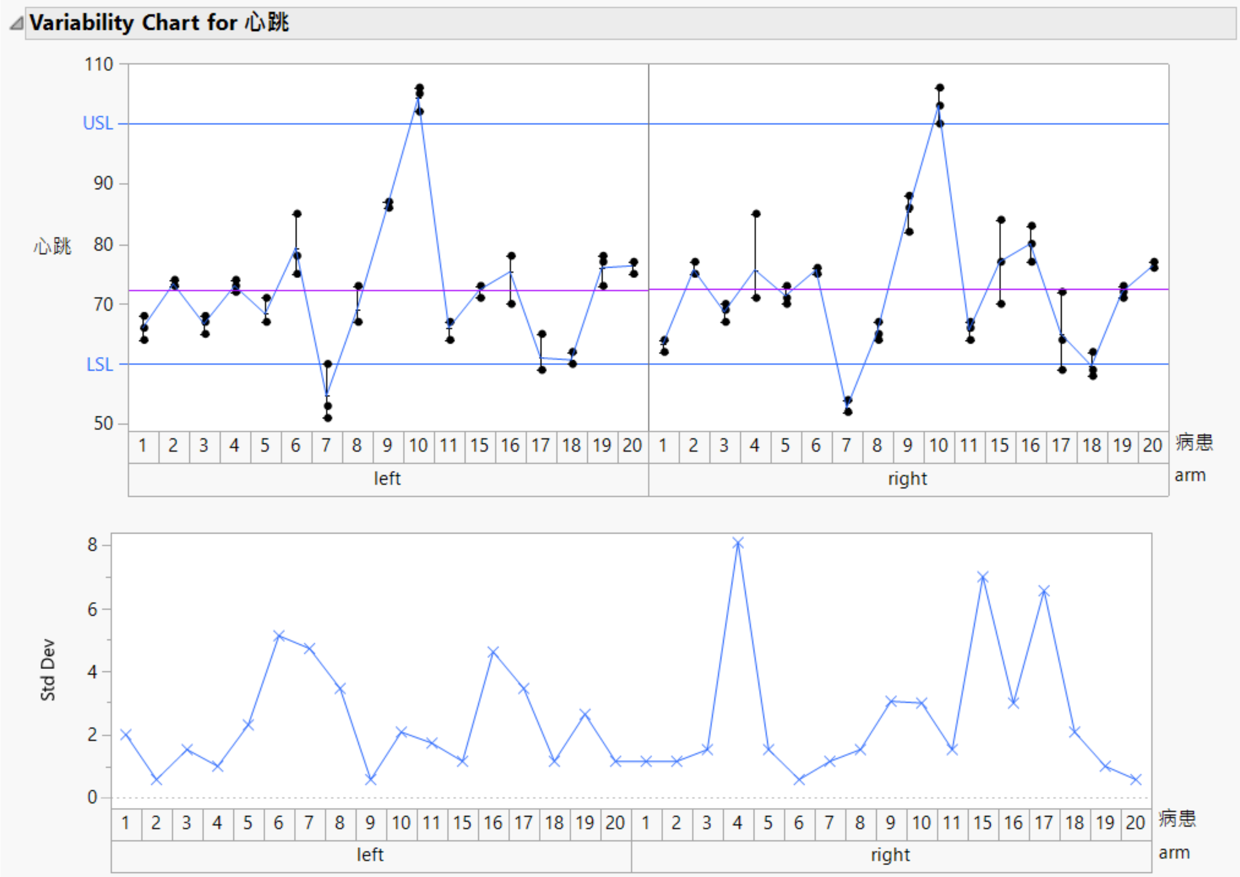

Figure 3. Variability and Standard Deviation Charts of Heart Rate Measurements for Left and Right Arm Trials

According to the Heart Rate Variability Chart, differences can be observed among the three repeated measurements of each sample, along with their corresponding standard deviations. Several data points exhibit relatively high standard deviation values, and considering that the repeatability exceeds 30%, the results indicate instability in the measurements. Focusing on potential misclassification, the error-prone samples demonstrate varying levels of risk. For Samples 4, 6, and 15, although the three measurements show considerable variation, the values remain within the middle range, resulting in a relatively low likelihood of being misclassified into hypertensive or hypotensive categories. In contrast, Samples 7 and 17 not only display large variation among the three measurements but also have values that are close to or exceed the threshold between high and low blood pressure. Given the repeatability error of more than 30%, these samples present a higher risk of misclassification by the device, potentially leading to delayed treatment or inappropriate allocation of medical resources. For the remaining samples, measurements generally exhibit small standard deviations and values within the normal range between hypertension and hypotension. Although some measurement error may still exist, the risk of misclassification is the lowest among all categories.

4.2 The Analysis of Hypertension

This study conducted a distribution analysis of heart-related measurements, with a particular focus on heart rate, to evaluate measurement stability.

Based on the clinically accepted tolerance range for heart rate (60–100 bpm), the Heart Rate Variability Chart indicated noticeable differences across the three repeated measurements of each sample, with some samples exhibiting relatively high standard deviations. The majority of measurements fell within the normal distribution range, suggesting that the device can provide reliable readings in most cases. However, Samples 4, 6, and 15, while showing greater internal variation, remained within the mid-range and therefore carried a relatively low risk of being misclassified as hypertensive or hypotensive. In contrast, Samples 7 and 17 not only demonstrated significant variability among repeated measurements but also presented values close to or exceeding critical thresholds, thereby posing a higher risk of misclassification. Overall, the distribution of heart rate measurements across samples largely aligned with clinical expectations, but the analysis revealed that high repeatability error significantly influenced the results. This indicates that device stability, rather than human physiological variation, is the primary factor affecting the accuracy of heart measurement distribution.

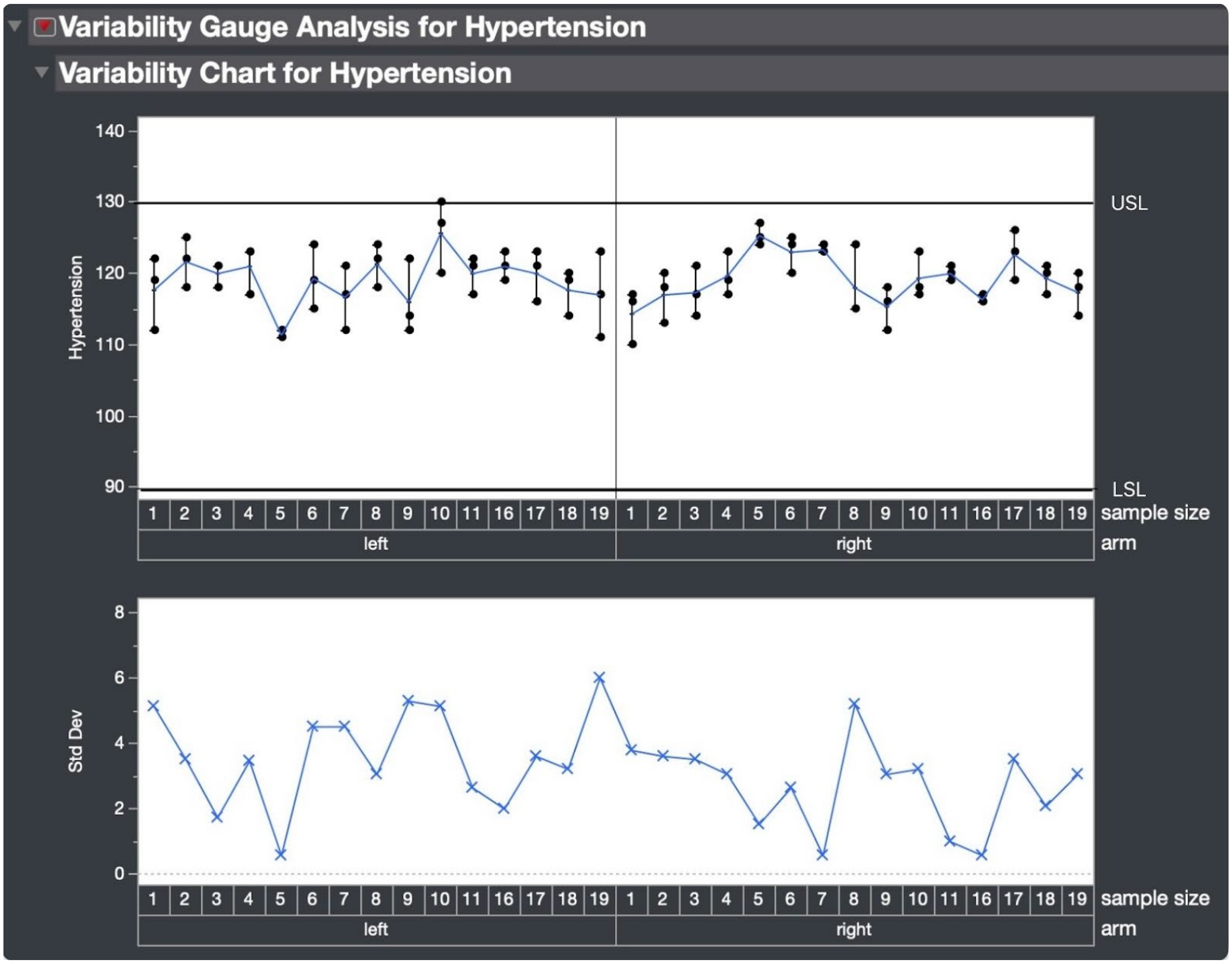

Figure 4. Variability Analysis Chart for Hypertension Measurements.

This Figure 4 shows the systolic blood pressure measurements of each participant for both the left and right arms.

According to the blood pressure measurement chart, most participants exhibited relatively large standard deviations in their blood pressure readings, indicating limited measurement stability. However, all recorded values remained within the clinically acceptable range, suggesting that the majority of participants did not show signs of hypertension.

It is noteworthy that, prior to testing, Sample No. 5 had been identified as a hypertensive individual, while Sample No. 16 exhibited below-average blood pressure levels. This finding suggests that the device may demonstrate reduced stability and accuracy when measuring high-risk groups, such as individuals with hypertension or hypotension.

Clinically, the acceptable range for systolic blood pressure is between 90 and 130 mmHg. According to the analysis chart, the standard deviations of the three consecutive measurements were relatively high, indicating limited repeatability and measurement stability of the device. This suggests that the instrument may not consistently provide precise results under repeated testing conditions, which could affect its overall reliability in long-term monitoring applications.

Although the standard deviations of the measurements were relatively high, the systolic blood pressure values of the participants were mostly concentrated between 90 and 130 mmHg, indicating that the overall distribution remained within the clinically acceptable range. This suggests that the general measurement trend was consistent and did not show significant deviation from expected values. However, due to the limited measurement stability, potential misjudgments may still occur among high-risk populations, such as older adults or individuals with cardiovascular diseases, as well as among those whose systolic pressures are close to the clinical threshold. To enhance accuracy and reliability, it is recommended to perform multiple repeated measurements or verify the results using medical-grade instruments.

4.3 The Analysis of Hypotension

Among the participants, sample size 2, 3, 7, 8, and 10 showed relatively large standard deviations in the measurements of the left hand, although all data remained within the normal range. In addition, sample size 9's left-hand measurement and sample size 1's right-hand measurement each showed one abnormal value among the three trials. For the right-hand measurements, sample size 1, 2, 3, 6, 11, 16, 17, and 18 also exhibited relatively high standard deviations. Except for sample size 1, whose one measurement result was abnormal, the data of all other participants were within the normal range.

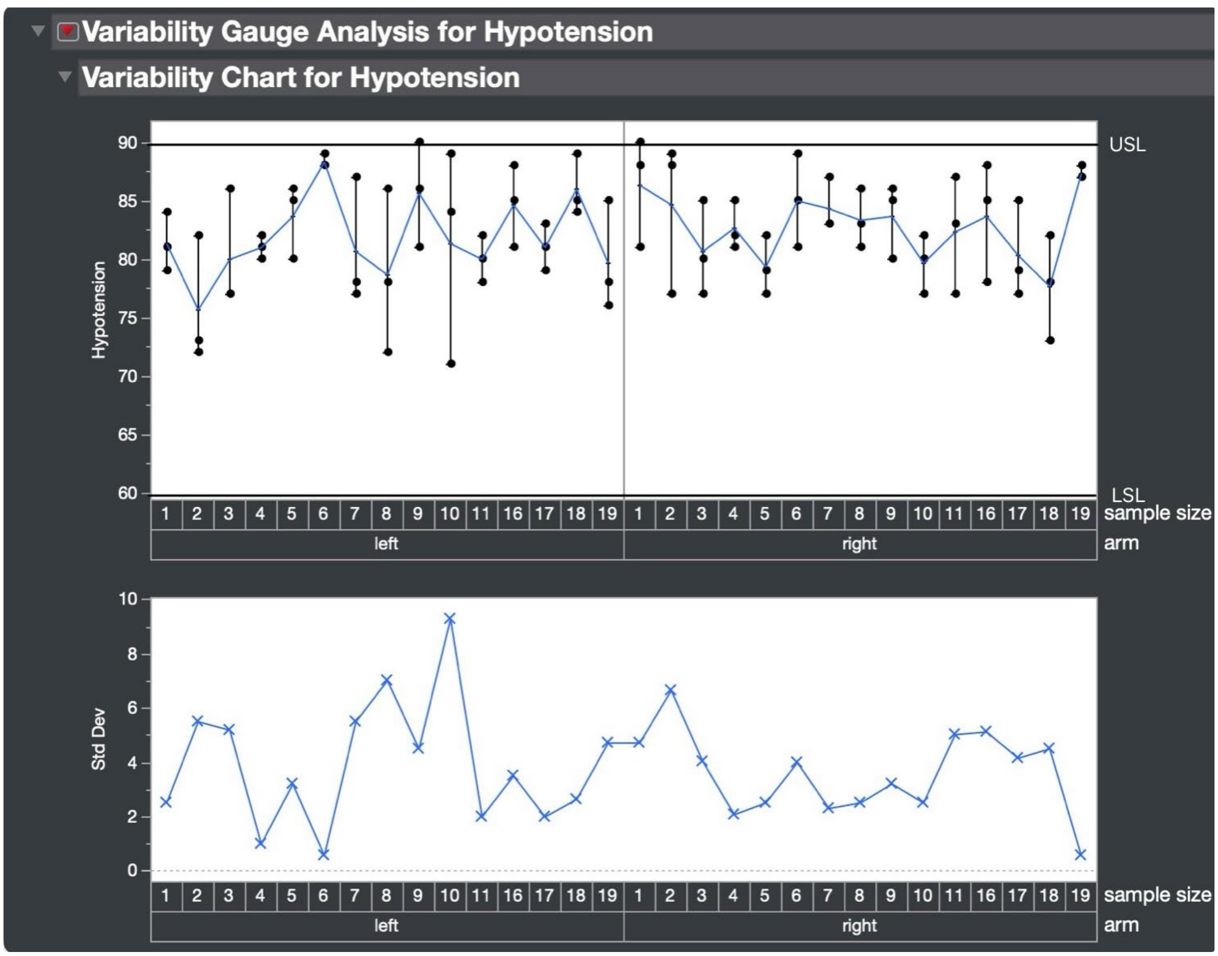

Figure 5. Variability Analysis Chart for Hypotension Measurements.

This Figure 5 shows the systolic blood pressure measurements of each participant for both the left and right arms.

5. Discussion

This study conducted a distribution analysis of heart-related measurements, with a particular focus on heart rate, to evaluate measurement stability.

Based on the clinically accepted tolerance range for heart rate (60–100 bpm), the Heart Rate Variability Chart indicated noticeable differences across the three repeated measurements of each sample, with some samples exhibiting relatively high standard deviations. The majority of measurements fell within the normal distribution range, suggesting that the device can provide reliable readings in most cases. However, Samples 4, 6, and 15, while showing greater internal variation, remained within the mid-range and therefore carried a relatively low risk of being misclassified as hypertensive or hypotensive. In contrast, Samples 7 and 17 not only demonstrated significant variability among repeated measurements but also presented values close to or exceeding critical thresholds, thereby posing a higher risk of misclassification. Overall, the distribution of heart rate measurements across samples largely aligned with clinical expectations, but the analysis revealed that high repeatability error significantly influenced the results. This indicates that device stability, rather than human physiological variation, is the primary factor affecting the accuracy of heart measurement distribution.

According to the low-pressure data, the overall standard deviations were generally high, and some participants showed both normal and abnormal results across trials. This indicates that the measurement stability still requires improvement. Such variability may be related to individual differences, measurement conditions, or the sensitivity of the instruments, and further investigation into these potential influencing factors is recommended.

Based on the results of this experiment, the stability of the device still requires improvement. When the measured values approach the threshold of the normal range, the system tends to produce misjudgments. In addition, the measurements are highly susceptible to external environmental influences, resulting in data fluctuations and instability. This issue is particularly evident during low-pressure measurements. Such instability may lead to misinterpretation in clinical applications, potentially causing delays in treatment or unnecessary consumption of medical resources.

Acknowledgements

The authors would like to thank the following personnel for their assistance in this study: Charles Chen, Yu Hua Cheng, Dai Lien-Guo and Yu Cheng Tai for advice in experimental design, statistical analysis, and revising the manuscript.

References

Automotive Industry Action Group (AIAG), *Measurement Systems Analysis Reference Manual*, 3rd ed., Troy, MI: AIAG, 2002.

Clark, C. E., Taylor, R. S., Shore, A. C., Ukoumunne, O. C., & Campbell, J. L., Association of a difference in systolic blood pressure between arms with vascular disease and mortality: A systematic review and meta-analysis, *The Lancet*, vol. 387, no. 10022, pp. 905–914, 2016.

Franklin, S. S., Gustin, W., Wong, N. D., Larson, M. G., Weber, M. A., Kannel, W. B., & Levy, D., Hemodynamic patterns of age-related changes in blood pressure: The Framingham Heart Study, *Circulation*, vol. 96, no. 1, pp. 308–315, 1997.

Freeman, R., Wieling, W., Axelrod, F. B., Benditt, D. G., Benarroch, E., Biaggioni, I., … & van Dijk, J. G., Consensus statement on the definition of orthostatic hypotension, neurally mediated syncope and the postural tachycardia syndrome, *Autonomic Neuroscience*, vol. 161, no. 1-2, pp. 46–48, 2011.

Grassi, G., Mark, A., & Esler, M., The sympathetic nervous system alterations in human hypertension, *Circulation Research*, vol. 116, no. 6, pp. 976–990, 2015.

Kim, B., & Baek, J., Photoplethysmography in Wearable Devices: A Comprehensive Review of Technological Advances, Current Challenges, and Future Directions, *Electronics*, vol. 12, no. 13, 2023.

Lee, J., Matsumura, K., Yamakoshi, K., Rolfe, P., Tanaka, S., & Yamakoshi, T., Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion, *IEEE*, 2017.

Li, J., Ma, J., Omisore, M., Liu, Y., Tang, H., Ao, P., Yan, Y., Wang, L., & Nie, Z., Noninvasive blood glucose monitoring using spatiotemporal ECG and PPG feature fusion and weight-based Choquet integral multimodal approach, *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2023.

Přibil, J., Přibilová, A., & Frollo, I., Comparative Measurement of the PPG Signal on Different Human Body Positions by Sensors Working in Reflection and Transmission Modes, *Engineering Proceedings*, pp. 69, 2020.

Sun, Y., & Thakor, N., Photoplethysmography Revisited: From Contact to Noncontact, From Point to Imaging, *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 463–477, 2016.

Van den Hoogen, P. C., Feskens, E. J., Nagelkerke, N. J., Menotti, A., Nissinen, A., & Kromhout, D., The relation between blood pressure and mortality due to coronary heart disease among men in different parts of the world, *New England Journal of Medicine*, vol. 342, no. 1, pp. 1–8, 2000.

Woodward, M., Webster, R., Murakami, Y., Barzi, F., Lam, T. H., Fang, X., … & Asia Pacific Cohort Studies Collaboration, The association between resting heart rate, cardiovascular disease and mortality: Evidence from 112,680 men and women in 12 cohorts, *European Journal of Preventive Cardiology*, vol. 21, no. 6, pp. 719–726, 2014.

Zhang, D., Wang, W., Li, F., Yu, J., Kang, X., Ma, Y., … & Li, G., Resting heart rate and all-cause and cardiovascular mortality in the general population: A meta-analysis, *Canadian Medical Association Journal*, vol. 188, no. 3, pp. E53–E63, 2016.

Biographies

Zan Chou is a junior year high school student with a long-standing passion for biology and physical sciences at Huaxing High School, he has actively engaged in experimental work since junior high school. His recent exposure to research involving medical instrumentation has deepened his interest in biomedical technologies. To support his experimental analysis, he has begun learning JMP statistical software and applying it to process and interpret physiological data.

Ni-Hsi Yeh is currently a grade 12 student at Fanghe Experimental High School. From a young age, she has shown strong ambition and passion for social issues and advocacy. Driven by a curiosity to explore societal needs and initiate action, she has participated in short-term mission and service teams in Kenya, Northern Thailand, and Taiwan from elementary through high school.In her junior year, she competed in the 2025 Civic Action Project Competition organized by the Legal Reform Education Foundation (LREF), earning Third Place and the "Most Popular Award." With a keen interest in addressing population aging and applying technology, she aims to learn more about data analytics and to develop innovative solutions. Through competition participation and presentations, she seeks to broaden her global perspective, cultivate social insight, and objectively contribute to positive change in society.

Pin-Jen Lai is currently a junior year high school student at Kang Chiao International School Linkou Campus. She is currently interested in pursuing a future in biological and data sciences, with a particular focus on translational research. She has participated in the 2024 iGEM competition to collectively create an optimized biosensor for atherosclerotic plaque. Recognizing the importance of preventive healthcare, she has continued to explore interdisciplinary approaches integrating biotechnology and statistical modeling to support early detection and diagnosis.