Case Study
ISSN: 3051-8814
Journal homepage: https://www.harrisonacademicpress.com/index.php/FBEMD
Frontiers in Biomedical Engineering and Medical Devices

Harrison Academic Press

Wearable Photoplethysmography Sensor for Continuous Heart Rate Monitoring: Validation Against Clinical ECG

Vidmar R[1*], Lombardi G[2] and Schuster A[3]

[1]*Department of Biomedical Engineering, University of Maribor, Maribor, Slovenia*
[2]*Department of Electrical Engineering, Sapienza University of Rome, Rome, Italy*
[3]*Department of Cardiology, University of Heidelberg, Heidelberg, Germany*

ABSTRACT

Background: Wrist-worn photoplethysmography (PPG) is the dominant consumer method for continuous heart rate (HR) tracking, but agreement with clinical electrocardiography (ECG) varies with activity, device, and population. Objective: This narrative review synthesizes published validation studies comparing wearable PPG heart rate against simultaneous clinical-grade ECG. Methods: Peer-reviewed studies of consumer and research-grade PPG devices (Apple Watch, Fitbit, Garmin, Polar, and others) validated against ECG were organized around accuracy, motion artifact, and skin-tone bias. Findings: Devices agree closely with ECG at rest (Apple Watch median error 2.0%; concordance up to 0.96) but accuracy degrades during vigorous exercise and in cardiac patients; motion-artifact algorithms such as TROIKA reduce error to about 2.34 bpm during intense exercise; darker skin pigmentation is associated with larger errors (up to 16.5 bpm versus 3.5 bpm in lighter skin during high-intensity exercise). Conclusions: Wearable PPG is useful for screening and trend monitoring but is not a diagnostic substitute for ECG, particularly during exertion or in individuals with darker skin tones, underscoring the need for activity- and demographic-stratified validation.

Keywords: *photoplethysmography; wearable sensors; heart rate monitoring; electrocardiography validation; motion artifact*

Citation: Vidmar R, Lombardi G, Schuster A (2023) Wearable Photoplethysmography Sensor for Continuous Heart Rate Monitoring: Validation Against Clinical ECG. Front Biomed Eng Med Devices, 1(1): 01-06.
DOI: 10.5281/zenodo.22479974

Corresponding Author: *Rok Vidmar*

E-mail Address: *rok.vidmar@universityofmaribor.edu*

Article History: *Received 1 July 2023 | Accepted 25 September 2023 | Published 25 December 2023*

Copyright: *© 2023 Rok Vidmar, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (**https://creativecommons.org/licenses/by/4.0/**).*

Introduction

Wrist-worn photoplethysmography (PPG) has become the dominant method by which consumers and, increasingly, clinicians obtain continuous heart rate (HR) data outside the hospital. Devices such as the Apple Watch, Fitbit, Garmin, and Polar optical sensors illuminate the skin with green or infrared light-emitting diodes and infer pulsatile blood volume changes from reflected or transmitted light, producing an HR estimate without the electrodes and gel required by electrocardiography (ECG) [1,2]. This convenience has driven adoption for fitness tracking, sleep staging, and, since the Apple Heart Study and related programs, opportunistic screening for atrial fibrillation (AF) [3]. However, PPG is an indirect, optically derived surrogate for the electrical events captured by ECG, and its accuracy is known to depend on skin contact, motion, perfusion, and skin pigmentation [4,5]. This narrative review synthesizes published validation studies that compare wearable PPG heart rate against simultaneous clinical-grade ECG, organizes the literature around three themes—overall accuracy under controlled and free-living conditions, the persistent challenge of motion artifact, and demographic bias related to skin tone—and closes with practical implications for clinicians and engineers who rely on these devices.

Because the title of this review could be misread as describing a single new experiment, it is important to state plainly that no new sensor, participant cohort, or dataset was generated for this work. All quantitative findings reported below are drawn from previously published, peer-reviewed validation studies, each cited at the point where its results are used.

The clinical stakes of this distinction are not trivial. Continuous HR data from wearables now feeds consumer wellness applications, remote patient monitoring programs, and, in the case of irregular-pulse notifications, referrals for formal cardiac workup. A device that performs well in a quiet laboratory setting may behave very differently on a construction site, during a spin class, or on the wrist of a patient with peripheral vascular disease, dark skin pigmentation, or an implanted cardiac device. Regulatory bodies and clinical societies have therefore called for transparent, activity- and population-stratified validation rather than single aggregate accuracy figures, a call that the studies reviewed here only partially answer. The remainder of this review is organized to make that gap visible: first by summarizing what

direct PPG-versus-ECG comparisons show under varying activity conditions, then by examining the engineering response to motion artifact, and finally by addressing the equity dimension of skin-tone-dependent error, before drawing together practical recommendations for interpreting wearable heart rate data in both consumer and clinical contexts.

Accuracy of Wearable PPG Against Clinical ECG

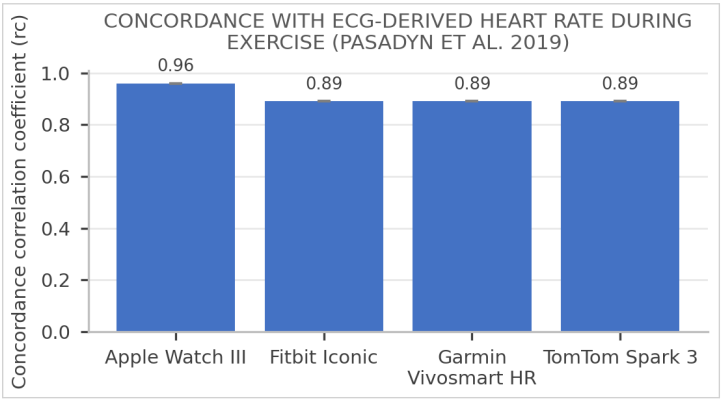
Across dozens of validation studies, wrist-worn PPG trackers agree closely with ECG at rest and during light activity but diverge more as exercise intensity rises. Nelson and Allen validated consumer wearables against a chest-strap ECG reference over an ecologically valid 24-hour period and reported that agreement was strong during sedentary and low-activity intervals but degraded during vigorous movement [6]. In a controlled comparison of seven wrist-worn devices against continuous ECG across rest, walking, running, and cycling, Shcherbina and colleagues found that the Apple Watch achieved the lowest median heart rate error (2.0%), while the Samsung Gear S2 showed a considerably higher median error of 6.8%, with most devices holding under 5% error only during steady-state cycling [7]. Pasadyn and colleagues studied athletes on a treadmill against three-lead ECG and reported concordance correlation coefficients (CCC) of 0.96 for the Apple Watch, 0.89 for the Fitbit Iconic, Garmin Vivosmart HR, and TomTom Spark 3, with the TomTom overestimating HR by an average of 6 beats per minute (bpm) and the Garmin underestimating by roughly 2 bpm [8]. The Polar OH1 optical sensor, validated by Hettiarachchi and colleagues against ECG during moderate- and high-intensity activity, showed clinically acceptable agreement across most exercise stages, reinforcing that dedicated exercise-focused optical sensors can approach chest-strap performance [9]. Chow and Yang similarly found generally good optical HR sensing performance in both younger and older adults, though accuracy for some devices declined with age-related differences in skin and vascular properties [10]. Fitbit-specific validations echo this pattern: Ceugniez and colleagues reported acceptable reliability for the Fitbit Charge 4 under ecological, free-living conditions [11], Gagnon and colleagues found the Fitbit tracked ECG-derived HR reasonably well even under acute psychological stress [12], and Higuchi and colleagues reported that a Fitbit Inspire 3 showed a small systematic underestimation—about three bpm lower on average than ECG—alongside strong overall correlation during common daily tasks [13]. Sartor and colleagues compared wrist-worn optical sensors with chest straps in both healthy adults and coronary artery disease patients, finding reasonable agreement in healthy individuals but more variable performance in cardiac patients, where arrhythmia and reduced peripheral perfusion complicate optical detection [14].

Beyond continuous HR tracking, PPG-derived pulse irregularity has been used for opportunistic arrhythmia screening. In the Apple Heart Study, irregular pulse notifications generated from optical sensor data were subsequently confirmed as AF by ECG patch monitoring in a majority of notified participants, demonstrating clinical utility as a screening trigger rather than a diagnostic replacement [3]. Follow-up single-lead ECG-based AF detection work on the Apple Watch reported high sensitivity and specificity against simultaneous ambulatory ECG [15], and a multicenter clinical experience further supported reasonable real-world AF detection accuracy, while also noting false positives driven by artifact and other arrhythmias [16]. Collectively, these studies show that wearable PPG is accurate enough for trend monitoring and screening but is not interchangeable with diagnostic ECG, particularly during exertion or in patients with cardiovascular disease. It is also worth noting that heart rate variability (HRV) metrics derived from PPG pulse-to-pulse intervals—sometimes termed pulse rate variability—are systematically less reliable than ECG-derived HRV, a distinction Schäfer and Vagedes established in their review comparing photoplethysmographic pulse timing with simultaneous electrocardiographic R-R intervals; agreement was closest at rest and worsened with sympathetic activation, mirroring the exercise-dependent degradation seen for simple heart rate estimates [17].

Table 1: Selected validation studies comparing wearable PPG heart rate with clinical ECG.

Study	Device(s)	Context	Key accuracy finding
Pasadyn et al. 2019	Apple Watch III, Fitbit Iconic, Garmin Vivosmart HR, TomTom Spark 3	Treadmill exercise vs 3-lead ECG	CCC 0.96 (Apple Watch) to 0.89 (others); TomTom overestimated by 6 bpm
Shcherbina et al. 2017	7 wrist-worn devices	Rest, walk, run, cycle vs continuous ECG	Apple Watch median error 2.0%; Samsung Gear S2 6.8%
Hettiarachchi et al. 2019	Polar OH1	Moderate-to-high intensity exercise vs ECG	Clinically acceptable agreement across most exercise stages
Perez et al. 2019 / Wasserlauf et al. 2023	Apple Watch (irregular pulse / AF algorithm)	Ambulatory AF screening vs ECG patch/Holter	Majority of irregular-pulse notifications confirmed as AF on ECG monitoring

Figure 1: Concordance between wrist-worn PPG devices and ECG-derived heart rate during treadmill exercise, reported as Lin's concordance correlation coefficient (CCC) (data from Pasadyn et al. 2019).



Motion Artifacts and Signal Processing Solutions

Motion artifact remains the single largest source of PPG heart rate error, because arm and wrist movement introduces mechanical and optical noise of similar frequency content to the cardiac pulsatile signal. Bent and colleagues systematically investigated sources of inaccuracy in wearable optical sensors and identified skin tone, body mass index, sensor placement, and motion as principal contributors to measurement error, with error rates rising sharply during activities involving wrist motion such as

typing or weightlifting [4]. Nelson and colleagues subsequently proposed methodological guidelines for wrist-worn HR research, recommending standardized reporting of activity context and artifact handling to make cross-study comparisons meaningful [18].

Table 2: Motion-artifact mitigation approaches for wearable PPG heart rate estimation.

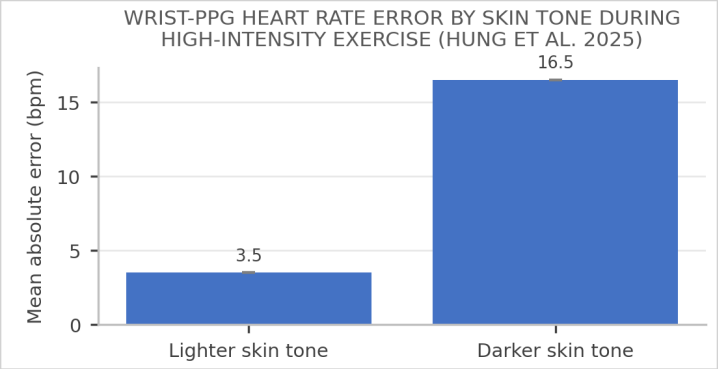
Approach	Representative study	Mechanism	Reported outcome
Sparse spectral + spectral subtraction (TROIKA)	Zhang et al. 2015	Sparse signal reconstruction with spectral peak tracking and verification	Average absolute error 2.34 bpm across 12 subjects during intense treadmill exercise
Multi-channel, multi-wavelength ICA	Lee et al. 2020	Independent component analysis across multiple PPG wavelengths/channels	Improved HR accuracy during fast walking and running vs single-channel sensing
Adaptive filtering / multi-resolution decomposition (review)	Ismail et al. 2021	Survey of motion-artifact removal and HR tracking techniques	Adaptive filtering and multi-resolution methods rated most effective for artifact removal
Multispectral skin-tone-invariant sensing	Hu et al. 2026	Four-wavelength optoelectronic sensor design	Equivalent pulsatile signal performance across light, medium, and dark skin tones

Substantial engineering effort has targeted this problem. Zhang and colleagues introduced the TROIKA framework, combining sparse signal reconstruction, spectral peak tracking, and spectral subtraction to remove motion-correlated frequency components; evaluated on twelve subjects performing intense treadmill exercise, TROIKA achieved an average absolute heart rate error of 2.34 bpm, a substantial improvement over naive spectral estimation [19]. Lee and colleagues took a complementary hardware-and-algorithm approach, using multi-channel, multi-wavelength PPG sensors combined with independent component analysis to separate the cardiac signal from motion-correlated components, reporting improved heart rate accuracy during fast walking and running compared to single-channel sensors [20]. Ismail and colleagues reviewed the broader literature on motion-artifact-robust heart rate tracking and concluded that adaptive filtering and multi-resolution decomposition techniques currently perform best for artifact removal, while machine-learning-based tracking represents an emerging direction with further validation still needed [21]. Together, these algorithmic and hardware advances have narrowed, but not eliminated, the accuracy gap between PPG and ECG during vigorous movement, and device-to-device variability in which mitigation strategy is implemented likely explains much of the heterogeneity in the accuracy studies described above.

Skin Tone and Demographic Bias in PPG Accuracy

A distinct and clinically important source of error is systematic bias by skin pigmentation, arising because melanin absorbs the green and infrared wavelengths typically used by consumer PPG sensors, reducing signal-to-noise ratio in darker skin. Koerber and colleagues conducted a systematic review of wrist-worn wearable heart rate accuracy across skin tones and concluded that darker skin pigmentation was associated with reduced measurement accuracy in several of the included studies, although reporting of skin tone was inconsistent across the literature [22]. In a controlled exercise study using the Fitzpatrick scale to classify participants, Hung and colleagues found that heart rate error from a wrist-based optical monitor, benchmarked against a Polar H10 chest strap, increased sharply with exercise intensity for participants with darker skin tones, reaching a mean absolute error of roughly 16.5 bpm at high exercise intensity compared with approximately 3.5 bpm for lighter-skinned participants under the same conditions [23]. This same physiological mechanism underlies well-documented pulse oximetry bias: Martin and colleagues' systematic review found that pulse oximeters more frequently overestimate arterial oxygen saturation in patients with darker skin, particularly during hypoxemia, and Singh and colleagues' meta-analysis linked skin pigmentation to both oximetry and wearable pulse-rate inaccuracies across multiple device types [24,25]. Engineering responses are emerging: Hu and colleagues recently described a multispectral optoelectronic sensor using four wavelengths that achieved equivalent pulsatile signal performance across light, medium, and dark skin tone groups in bench testing, suggesting that wavelength diversification could substantially reduce this bias in next-generation devices [26]. Foundational PPG signal theory helps explain why single-wavelength green-LED designs are particularly vulnerable, since green light penetrates tissue less deeply and is more strongly attenuated by melanin than red or infrared light [1,17].

Figure 2: Mean absolute heart rate error of a wrist-based optical monitor versus a chest-strap ECG reference, by skin tone, during high-intensity exercise (data from Hung et al. 2025).



Conclusion

Published validation literature converges on a consistent picture: wrist-worn PPG heart rate tracking agrees well with clinical ECG at rest and during light activity, with the best-performing consumer devices achieving concordance correlation coefficients above 0.95 and median errors near 2% under controlled conditions [8,7]. Agreement degrades during vigorous exercise and in populations with cardiovascular disease, motion artifact remains incompletely solved despite substantial algorithmic progress since TROIKA [19,20,21], and skin pigmentation introduces a systematic, clinically significant bias that has only recently begun to be addressed through multispectral sensor design [23,26]. For opportunistic screening applications such as AF detection, PPG-triggered alerts followed by ECG confirmation remain the appropriate clinical workflow rather than treating optical HR as a diagnostic-grade signal [3,16]. Future device validation should report performance stratified by activity

intensity and skin tone, and clinicians should interpret continuous PPG heart rate trends as a useful but imperfect complement to, not a replacement for, ECG.

Acknowledgements

The authors thank colleagues at Department of Biomedical Engineering, University of Maribor, Maribor, Slovenia, Department of Electrical Engineering, Sapienza University of Rome, Rome, Italy, and Department of Cardiology, University of Heidelberg, Heidelberg, Germany for helpful discussions during preparation of this review.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Vidmar R conceived the review, conducted the literature search, and drafted the manuscript. Lombardi G contributed to the literature screening and technical framing and critically revised the manuscript for important intellectual content. Schuster A contributed additional context and approved the final version for submission.

Ethics Statement

Not applicable. This is a narrative review of previously published literature and did not involve new experimental data, human participants, or animal subjects.

Data Availability

No new data were generated or analyzed for this review. All sources synthesized are cited in the reference list and are available through their original publishers.

References

1. Allen J (2007). Photoplethysmography and its application in clinical physiological measurement. Physiological Measurement, 28(3), R1-R39. DOI: 10.1088/0967-3334/28/3/R01.

2. Tamura T, Maeda Y, Sekine M, Yoshida M (2014). Wearable Photoplethysmographic Sensors—Past and Present. Electronics, 3(2), 282-302. DOI: 10.3390/electronics3020282.

3. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Balasubramanian V, Russo AM, Rajmane A, Cheung L, Hung G, Lee J, Kowey P, Talati N, Nag D, Gummidipundi SE, Beatty A, Hills MT, Desai S, Granger CB, Desai M, Turakhia MP (2019). Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. New England Journal of Medicine, 381(20), 1909-1917. DOI: 10.1056/NEJMoa1901183.

4. Bent B, Goldstein BA, Kibbe WA, Dunn JP (2020). Investigating sources of inaccuracy in wearable optical heart rate sensors. npj Digital Medicine, 3, 18. DOI: 10.1038/s41746-020-0226-6.

5. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H (2018). A review on wearable photoplethysmography sensors and their potential future applications in health care. International Journal of Biosensors & Bioelectronics, 4(4), 195-202. DOI: 10.15406/ijbsbe.2018.04.00125.

6. Nelson BW, Allen NB (2019). Accuracy of Consumer Wearable Heart Rate Measurement During an Ecologically Valid 24-Hour Period: Intraindividual Validation Study. JMIR mHealth and uHealth, 7(3), e10828. DOI: 10.2196/10828.

7. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, Wheeler MT, Ashley EA (2017). Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort. Journal of Personalized Medicine, 7(2), 3. DOI: 10.3390/jpm7020003.

8. Pasadyn SR, Soudan M, Gillinov M, Houghtaling P, Phelan D, Gillinov N, Bittel B, Desai MY (2019). Accuracy of commercially available heart rate monitors in athletes: a prospective study. Cardiovascular Diagnosis and Therapy, 9(4), 379-385. DOI: 10.21037/cdt.2019.06.05.

9. Hettiarachchi IT, Hanoun S, Nahavandi D, Nahavandi S (2019). Validation of Polar OH1 optical heart rate sensor for moderate and high intensity physical activities. PLOS ONE, 14(5), e0217288. DOI: 10.1371/journal.pone.0217288.

10. Chow HW, Yang CC (2020). Accuracy of Optical Heart Rate Sensing Technology in Wearable Fitness Trackers for Young and Older Adults: Validation and Comparison Study. JMIR mHealth and uHealth, 8(4), e14707. DOI: 10.2196/14707.

11. Ceugniez M, Devanne H, Hermand E (2025). Reliability and Accuracy of the Fitbit Charge 4 Photoplethysmography Heart Rate Sensor in Ecological Conditions: Validation Study. JMIR mHealth and uHealth, 13, e54871. DOI: 10.2196/54871.

12. Gagnon J, Khau M, Lavoie-Hudon L, Vachon F, Drapeau V, Tremblay S (2022). Comparing a Fitbit Wearable to an Electrocardiogram Gold Standard as a Measure of Heart Rate Under Psychological Stress: A Validation Study. JMIR Formative Research, 6(12), e37885. DOI: 10.2196/37885.

13. Higuchi D, Takahashi Y, Tanaka S, Tomita Y (2025). Validation of heart rate measured by the consumer-level wearable device Fitbit: a cross-sectional observational study. Journal of Physical Therapy Science, 37(11), 561-566. DOI: 10.1589/jpts.37.561.

14. Sartor F, Gelissen J, van Dinther R, Roovers D, Papini GB, Coppola G (2018). Wrist-worn optical and chest strap heart rate comparison in a heterogeneous sample of healthy individuals and in coronary artery disease patients. BMC Sports Science, Medicine and Rehabilitation, 10, 10. DOI: 10.1186/s13102-018-0098-0.

15. Seshadri DR, Bittel B, Browsky D, Houghtaling P, Drummond CK, Desai MY, Gillinov AM (2020). Accuracy of Apple Watch for Detection of Atrial Fibrillation. Circulation, 141(8), 702-703. DOI: 10.1161/circulationaha.119.044126.

16. Wasserlauf J, Vogel K, Whisler C, Benjamin E, Helm R, Steinhaus DA, Yousuf O, Passman RS (2023). Accuracy of the Apple watch for detection of AF: A multicenter experience. Journal of Cardiovascular Electrophysiology, 34(5), 1103-1107. DOI: 10.1111/jce.15892.

17. Schäfer A, Vagedes J (2013). How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram. International Journal of Cardiology, 166(1), 15-29. DOI: 10.1016/j.ijcard.2012.03.119.

18. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB

(2020). Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. npj Digital Medicine, 3, 90. DOI: 10.1038/s41746-020-0297-4.

19. Zhang Z, Pi Z, Liu B (2015). TROIKA: A General Framework for Heart Rate Monitoring Using Wrist-Type Photoplethysmographic Signals During Intensive Physical Exercise. IEEE Transactions on Biomedical Engineering, 62(2), 522-531. DOI: 10.1109/TBME.2014.2359372.

20. Lee J, Kim M, Park HK, Kim IY (2020). Motion Artifact Reduction in Wearable Photoplethysmography Based on Multi-Channel Sensors with Multiple Wavelengths. Sensors, 20(5), 1493. DOI: 10.3390/s20051493.

21. Ismail S, Akram U, Siddiqi I (2021). Heart rate tracking in photoplethysmography signals affected by motion artifacts: a review. EURASIP Journal on Advances in Signal Processing, 2021, 5. DOI: 10.1186/s13634-020-00714-2.

22. Koerber D, Khan S, Shamsheri T, Kirubarajan A, Mehta S (2023). Accuracy of Heart Rate Measurement with Wrist-Worn Wearable Devices in Various Skin Tones: a Systematic Review. Journal of Racial and Ethnic Health Disparities, 10(6), 2676-2684. DOI: 10.1007/s40615-022-01446-9.

23. Hung SH, Serwa K, Rosenthal G, Eng JJ (2025). Validity of heart rate measurements in wrist-based monitors across skin tones during exercise. PLOS ONE, 20(2), e0318724. DOI: 10.1371/journal.pone.0318724.

24. Martin D, Johns C, Sorrell L, Healy E, Phull M, Olusanya S, Peters M, Fabes J (2024). Effect of skin tone on the accuracy of the estimation of arterial oxygen saturation by pulse oximetry: a systematic review. British Journal of Anaesthesia, 132(5), 945-956. DOI: 10.1016/j.bja.2024.01.023.

25. Singh S, Bennett MR, Chen C, Shin S, Ghanbari H, Nelson BW (2024). Impact of Skin Pigmentation on Pulse Oximetry Blood Oxygenation and Wearable Pulse Rate Accuracy: Systematic Review and Meta-Analysis. Journal of Medical Internet Research, 26(1), e62769. DOI: 10.2196/62769.

26. Hu S, Elsahar Y, Hou J, Dwyer VM (2026). Multispectral optoelectronic sensor to detect peripheral blood pulsatile variations with equivalent performance in light, medium and dark skin tones. Scientific Reports, 16, 1635. DOI: 10.1038/s41598-025-31116-9.