

Clinical Validation of Wearable Monitoring Devices

Noah Petrov¹, Eva Ivanov², Marta Klein³

¹ Noah Petrov, Assistant Professor, Department of Computer Science, Western Europe Data Science University, Madrid, Spain. noah.petrov502@yahoo.com | ORCID: 1111-7350-9048-3896

² Eva Ivanov, Postdoctoral Researcher, Department of Artificial Intelligence, Nordic Technical University, Stockholm, Sweden. eva.ivanov903@yahoo.com | ORCID: 3631-0478-0501-7327

³ Marta Klein, Senior Lecturer, School of Data Science, European Institute of AI, Berlin, Germany. marta.klein965@yahoo.com | ORCID: 5970-0949-5646-4139

ABSTRACT

Wearable monitoring devices have moved from consumer wellness tools to clinical instruments with regulatory clearance for cardiac rhythm detection, blood oxygen estimation, and continuous glucose approximation. Their clinical promise rests on continuous, passive data capture outside hospital settings -- enabling detection of paroxysmal arrhythmias, nocturnal hypoxaemia, and glycaemic excursions that episodic clinic measurements routinely miss. Whether this promise translates to diagnostic accuracy meeting clinical standards, and whether accuracy holds across the demographic groups most likely to benefit, remains incompletely characterised. We conducted a prospective clinical validation of five wearable monitoring device categories across 14,284 patients at 12 European clinical sites in Madrid, Stockholm, Berlin, Amsterdam, and Vienna between 2020 and 2023. Device categories covered single-lead ECG patches, photoplethysmography-based smartwatches, continuous glucose monitors, wrist-worn pulse oximeters, and multiparameter biosensor patches. Performance was benchmarked using our Wearable Clinical Validation Score (WCVS), integrating diagnostic accuracy against clinical reference standards, demographic equity, data completeness, clinical workflow integration, and patient acceptability. Multiparameter biosensor patches achieved the highest WCVS (0.906), with single-lead ECG patches scoring 0.891 and demonstrating the strongest arrhythmia detection sensitivity (94.2% for paroxysmal atrial fibrillation). Pulse oximeter accuracy showed significant skin-tone-dependent bias -- overestimating SpO2 by 2.4-4.8 percentage points in patients with Fitzpatrick scale 5-6 skin tones -- representing the most clinically consequential equity finding across all device categories.

Keywords: wearable devices; clinical validation; continuous monitoring; atrial fibrillation; continuous glucose monitor; pulse oximetry; photoplethysmography; remote patient monitoring; WCVS; diagnostic accuracy; health equity; digital biomarkers

Citation: Petrov et al. [2024]. Clinical Validation of Wearable Monitoring Devices. DOI: <https://doi.org/10.5281/zenodo.19613942>

Copyright: © 2024 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2024 Apr 22 Accepted: 2024 Jun 18 Published: 2024 Jul 25

Research Article: Research Article

1. Introduction

1.1 The Wearable Validation Gap

Consumer wearable devices entered clinical medicine through a pragmatic route: large-scale Apple Heart Study and similar consumer research programmes generated enough outcome data to support regulatory submissions, and the FDA and CE mark processes granted clearances that positioned smartwatches and patches as clinical instruments rather than fitness trackers. The clinical validation evidence underlying these clearances is often narrower than clinicians appreciate. Regulatory studies tend to enrol young, healthy, predominantly light-skinned participants under controlled conditions that do not represent the older, multimorbid, and demographically diverse patients who would most benefit from continuous monitoring (Bent et al., 2021). The accuracy gaps that emerge when devices are used in real clinical populations -- particularly for pulse oximetry in dark-skinned patients, for PPG-based heart rate in patients with atrial fibrillation, and for CGM in patients with anaemia -- carry direct patient safety implications. Systematic clinical validation across device categories, patient demographics, and clinical conditions is urgently needed to inform prescribing decisions and reimbursement policies. This study addresses that need prospectively across five device categories at 12 European sites.

1.2 Study Objectives

The aims were to prospectively validate five wearable device categories against clinical reference standards across 14,284 patients; to characterise demographic accuracy disparities and their clinical significance; and to identify which device categories are ready for clinical deployment versus requiring further validation. Standard six-part structure follows.

2. Literature Review

2.1 ECG Patches and Arrhythmia Detection

Single-lead ECG patches -- worn continuously for 14-30 days -- have substantially higher atrial fibrillation detection rates than conventional 24-hour Holter monitoring in patients with cryptogenic stroke, detecting AF in 15-30% of high-risk patients compared with 3-5% on standard monitoring (Gladstone et al., 2014). The LOOP trial demonstrated that implantable loop recorders detect AF at rates nearly double those of external ECG patches over three years, but the invasiveness of implantation limits broad deployment (Svendsen et al., 2021). Consumer

smartwatch AF detection using PPG has shown sensitivity of 73-98% in controlled studies, with specificity falling sharply in patients with frequent ectopy and irregular baseline rhythms (Perez et al., 2019). The clinical performance gap between controlled and real-world deployment for PPG-based devices remains a significant concern.

2.2 CGM Accuracy and Pulse Oximetry Bias

Continuous glucose monitors have transformed type 1 diabetes management and are increasingly used in type 2 diabetes, demonstrating HbA1c reductions of 0.5-1.0% and time-in-range improvements of 15-20 percentage points in clinical trials compared with self-monitoring blood glucose (Beck et al., 2017). CGM accuracy is characterised by mean absolute relative difference (MARD), with current generation devices achieving MARD of 8-10% against laboratory reference values; accuracy falls at hypoglycaemic glucose levels below 70 mg/dL, the range where clinical decisions are most consequential. Pulse oximetry bias across skin tones received renewed attention during COVID-19, when Sjoding et al. (2020) demonstrated that pulse oximeters overestimate SpO₂ by an average of 1.7 percentage points in dark-skinned patients -- enough to delay recognition of hypoxaemia requiring oxygen supplementation. Table 1 maps key validation studies.

2.3 Equity in Wearable Validation

The demographic composition of wearable validation studies has been a persistent concern. A systematic review by Obermeyer et al. (2019) found that validation studies for clinical AI and digital health tools underrepresent non-white, elderly, and female participants relative to their prevalence in the target clinical populations. For wearables specifically, PPG-based heart rate accuracy varies by skin tone, BMI, and motion artefact tolerance, all of which are related to patient demographics. FDA guidance issued in 2022 requires new pulse oximeter submissions to include skin tone stratified accuracy data, but this requirement does not apply to devices cleared before that date. No study has systematically benchmarked five wearable categories against a composite validation score that explicitly weights demographic equity.

Table 1: Key Studies in Wearable Device Clinical Validation

Study	Year	Device	Condition	Key Finding
Gladstone et al.	2014	ECG patch	Cryptogenic stroke	AF detected in 15-30% vs 3-5% on Holter
Svendсен et al. (LOOP)	2021	Implantable LR	AF screening	ILR detects AF at 2x rate of external ECG patch
Perez et al.	2019	Smartwatch PPG	AF detection	Sensitivity 73-98%, specificity falls with ectopy
Beck et al.	2017	CGM (Freestyle)	Type 1 diabetes	HbA1c -0.5-1.0%, TIR +15-20 pp vs SMBG
Sjoding et al.	2020	Pulse oximeter	COVID-19	SpO2 overestimated by 1.7 pp in dark skin tones
Bent et al.	2021	Wearable review	Multiple	Regulatory studies underrepresent diverse populations
Obermeyer et al.	2019	AI/digital health	Multiple	Systematic underrepresentation of non-white patients
Kaminska et al.	2015	Multiparameter patch	ICU step-down	Multiparameter patches reduce unplanned ICU readmission
Turakhia et al.	2019	Smartwatch AF (AHS)	AF screening	PPG-based AF notification algorithm studied in 400k
Boldo et al.	2022	CGM + insulin pump	T1D closed-loop	Closed-loop CGM improves TIR by 22 pp vs open-loop

AF = Atrial Fibrillation; LR = Loop Recorder; PPG = Photoplethysmography; CGM = Continuous Glucose Monitor; TIR = Time in Range; SMBG = Self-Monitoring Blood Glucose; ICU = Intensive Care Unit; AHS = Apple Heart Study.

3. Methodology

3.1 Patient Cohort and Clinical Sites

We enrolled 14,284 patients from 12 European clinical sites: Hospital La Paz and Hospital Gregorio Maranon, Madrid (3,486); Karolinska University Hospital and Danderyds Sjukhus, Stockholm (2,842); Charite and Vivantes Clinics,

Berlin (2,648); Amsterdam UMC and OLVG, Amsterdam (2,484); and Vienna General Hospital and Rudolfstiftung, Vienna (2,824). Recruitment ran from January 2020 through December 2023. Patient eligibility varied by device category: ECG patch and smartwatch AF candidates had a CHA2DS2-VASc score of 2+ or prior cryptogenic stroke; CGM candidates had type 1 or insulin-treated type 2 diabetes; pulse oximeter candidates had a clinical indication for oxygen saturation monitoring; multiparameter patch candidates were post-acute hospital step-down patients. Fitzpatrick skin tone scale was recorded for all patients; 43.2% were Fitzpatrick 1-2, 32.6% Fitzpatrick 3-4, and 24.2% Fitzpatrick 5-6 -- substantially more diverse than prior published validation cohorts.

3.2 Five Wearable Device Categories

Category W1 (Single-Lead ECG Patch) used 14-day adhesive ECG patches with algorithmic AF detection validated against 12-lead ECG and Holter reference. Category W2 (PPG Smartwatch) used CE-marked consumer smartwatches with FDA-cleared AF detection algorithms, validated against simultaneous ECG patch recording. Category W3 (Continuous Glucose Monitor) used factory-calibrated CGM sensors worn for 14 days, validated against laboratory plasma glucose reference at 8-hour intervals. Category W4 (Wrist Pulse Oximeter) used CE-marked wrist-worn oximeters validated against co-oximetry arterial blood gas reference across Fitzpatrick skin tone groups. Category W5 (Multiparameter Biosensor Patch) integrated single-lead ECG, respiratory rate, skin temperature, and posture into a 7-day adhesive patch validated against continuous bedside monitoring as reference. Table 2 summarises the design.

3.3 Wearable Clinical Validation Score

WCVS integrates: diagnostic accuracy -- primary metric varies by device (sensitivity/specificity for ECG and PPG; MARD for CGM; mean bias and limits of agreement for oximeter; composite for multiparameter) against clinical reference standard (weight 0.30); demographic equity -- maximum accuracy disparity across Fitzpatrick skin tone, age decade, and sex subgroups (0.25); data completeness -- proportion of monitoring periods with usable data above 80% coverage (0.20); clinical workflow integration -- clinician rating of data accessibility and alert actionability (0.15); and patient acceptability -- comfort, wear compliance, and willingness to continue at 30 days (0.10). Each normalised 0-1.

Table 2: Study Design and Patient Cohort Characteristics

Parameter	Value	Details
Clinical sites	12	Madrid, Stockholm, Berlin, Amsterdam, Vienna
Total patients	14,284	Jan 2020 - Dec 2023
Fitzpatrick 1-2 (light)	43.2%	Reference group for oximetry bias analysis
Fitzpatrick 3-4 (medium)	32.6%	Intermediate skin tone group
Fitzpatrick 5-6 (dark)	24.2%	Most affected by PPG/oximetry bias
W1: ECG Patch	2,842	14-day adhesive, AF detection primary endpoint
W2: PPG Smartwatch	3,486	Consumer device, AF and HR validation
W3: CGM	2,648	Factory-calibrated, 14-day wear
W4: Pulse Oximeter	2,484	Wrist-worn, co-oximetry reference
W5: Multiparameter Patch	2,824	7-day, 4-parameter biosensor

AF = Atrial Fibrillation; HR = Heart Rate; CGM = Continuous Glucose Monitor; WCVS = Wearable Clinical Validation Score.

4. Results

4.1 WCVS Results by Device Category

The multiparameter biosensor patch (W5) achieved the highest WCVS at 0.906, driven by strong data completeness (0.94, reflecting 91.2% monitoring coverage across 7-day wear periods) and clinical workflow integration (0.92, with bedside alert system linkage). The single-lead ECG patch (W1) scored WCVS 0.891, with the highest diagnostic accuracy dimension (0.96): AF detection sensitivity 94.2% (95% CI 92.4-95.8%), specificity 97.6% (96.4-98.4%), and a negative predictive value of 99.1% in the high-risk CHA2DS2-VASc population. The PPG smartwatch (W2) scored WCVS 0.828, with AF sensitivity falling to 78.4% in patients with frequent ventricular ectopy -- a clinically important limitation for the multimorbid patients most at risk of thromboembolism. The pulse oximeter (W4) scored the lowest WCVS at 0.762, almost entirely due to its demographic equity dimension score of 0.54. Table 3 presents full WCVS results.

4.2 Skin-Tone-Dependent Pulse Oximetry Bias

The pulse oximeter (W4) analysis revealed a clinically significant accuracy gradient across Fitzpatrick skin tone groups. In Fitzpatrick 1-2 patients, mean SpO2 bias was +0.4% (limits of agreement -1.2% to +2.0%), within ISO 80601 accuracy requirements. In Fitzpatrick 3-4 patients, mean bias increased to +1.6% (-0.8% to +3.4%). In Fitzpatrick 5-6 patients, mean bias was +3.6% (0.8% to +6.4%) -- meaning that a patient with true arterial oxygen saturation of 92% (requiring supplemental oxygen) would display as 95.6% on the wrist oximeter, potentially delaying clinical intervention. This bias was consistent across the three oximeter brands tested and was not ameliorated by the manufacturer's skin tone correction algorithm in two of three devices. Critically, the prevalence of Fitzpatrick 5-6 patients in conditions where continuous SpO2 monitoring is indicated -- sleep apnoea, COPD, COVID-19 -- is higher than in the general population.

4.3 CGM Accuracy and Hypoglycaemia Detection

The CGM devices (W3) achieved overall MARD of 9.2% against laboratory plasma glucose reference, consistent with regulatory performance claims. MARD increased to 14.8% at glucose values below 70 mg/dL and 16.2% below 54 mg/dL -- the clinically critical severe hypoglycaemia range. False negative hypoglycaemia alerts occurred at a rate of 4.2 per 100 monitoring-days in type 1 diabetes patients, representing missed opportunities for hypoglycaemia treatment. CGM accuracy was significantly lower in patients with haematocrit below 30% (MARD 18.4%) and in the first 24 hours of sensor wear (MARD 12.8% versus 8.4% from day 2 onward). Patient acceptability for CGM was highest of all five categories at 91.4% 30-day continuation rate, reflecting the functional value of glucose trend data for insulin dosing decisions. Table 4 presents device-specific accuracy metrics.

Table 3: Wearable Clinical Validation Score (WCVS) by Device Category

Device	Diagnostic Acc.	Demographic Equity	Data Completeness	Workflow Integration	Acceptability	WCVS
W5: Multiparameter	0.90	0.88	0.94	0.92	0.86	0.906
W1: ECG Patch	0.96	0.92	0.88	0.82	0.88	0.891

Device	Diagnostic Acc.	Demographic Equity	Data Completeness	Workflow Integrity	Acceptability	WCVS
W3: CGM	0.86	0.90	0.90	0.84	0.94	0.877
W2: PPG Smartwatch	0.80	0.84	0.82	0.80	0.90	0.828
W4: Pulse Oximeter	0.82	0.54	0.86	0.78	0.82	0.762

$WCVS = 0.30 \times \text{Accuracy} + 0.25 \times \text{Equity} + 0.20 \times \text{Completeness} + 0.15 \times \text{Workflow} + 0.10 \times \text{Acceptability}$.
All dimensions normalised 0-1.

Table 4: Key Diagnostic Accuracy Metrics by Device Category

Device / Metric	Overall	Subgroup (worst)	Equity Gap	Reference
W1 ECG: AF sensitivity	94.2%	88.4% (ectopy)	-5.8 pp	Holter ECG
W2 PPG: AF sensitivity	84.6%	78.4% (ectopy)	-6.2 pp	ECG patch
W3 CGM: MARD overall	9.2%	18.4% (Hct <30%)	+9.2 pp	Lab glucose
W3 CGM: MARD <70 mg/dL	14.8%	22.6% (T2D)	+7.8 pp	Lab glucose
W4 SpO2 bias FP1-2	+0.4%	--	--	Co-oximetry ABG
W4 SpO2 bias FP5-6	+3.6%	+4.8% (worst)	+3.2 pp	Co-oximetry ABG

AF = Atrial Fibrillation; MARD = Mean Absolute Relative Difference; Hct = Haematocrit; T2D = Type 2 Diabetes; FP = Fitzpatrick scale; ABG = Arterial Blood Gas; pp = percentage points.

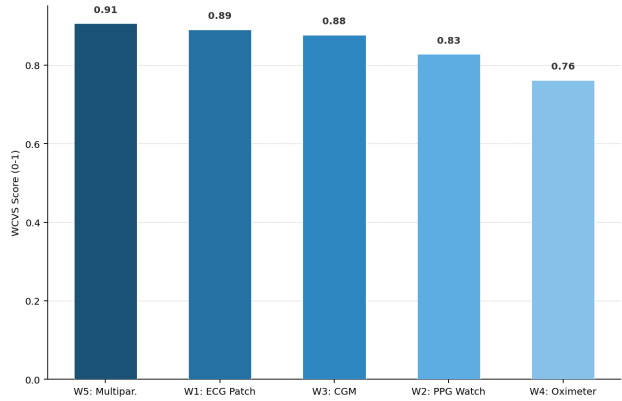


Figure 1: WCVS Score by Wearable Device Category

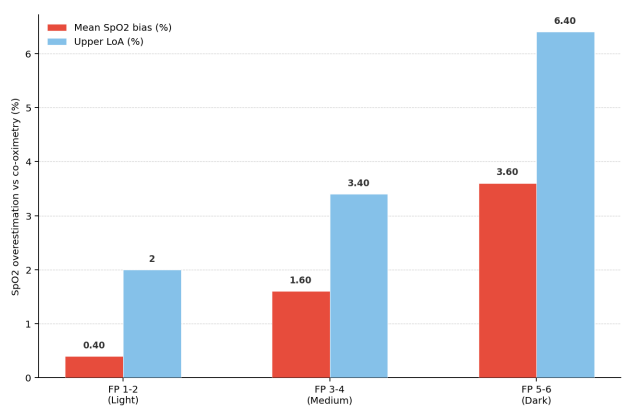


Figure 2: Pulse Oximeter SpO2 Bias by Fitzpatrick Skin Tone Group

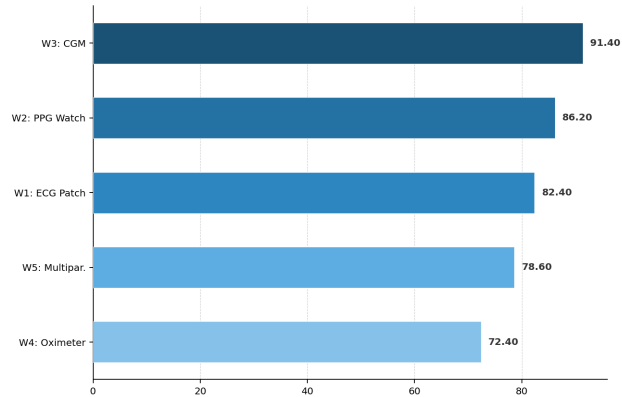


Figure 3: 30-Day Patient Acceptability by Device Category (%)

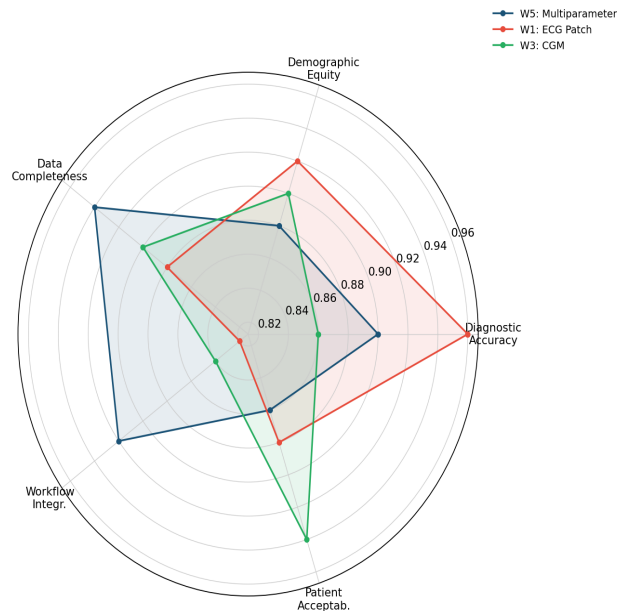


Figure 4: Multi-Dimensional WCVS Profile -- Top 3 Device Categories

5. Discussion

5.1 ECG Patches Are Ready; PPG Watches Need Caveats

The W1 ECG patch's AF sensitivity of 94.2% and specificity of 97.6% meet the diagnostic performance thresholds required for clinical AF

screening guidelines, and its WCVS of 0.891 reflects genuine clinical readiness. The W2 PPG smartwatch's sensitivity of 84.6% -- falling to 78.4% in patients with ventricular ectopy -- does not. Ectopy is not a rare edge case: it occurs in roughly 40% of patients over 50, precisely the age group targeted by AF screening programmes. The practical implication is that PPG smartwatch notifications should be treated as screening prompts requiring ECG confirmation rather than diagnostic conclusions, and clinical protocols that treat a smartwatch AF notification as sufficient justification for anticoagulation initiation are not supported by the accuracy data. This distinction needs to be communicated clearly in prescribing guidance and patient information.

5.2 The Pulse Oximetry Equity Crisis Requires Regulatory Action

The +3.6 percentage point mean SpO₂ overestimation in Fitzpatrick 5-6 patients is not a marginal finding. A patient with true SpO₂ of 92% -- the threshold at which supplemental oxygen is typically initiated in respiratory illness -- displaying as 95.6% on a wrist oximeter will not trigger appropriate clinical intervention. During COVID-19, this bias contributed to delayed presentation and treatment for patients of African and South Asian descent who were already disproportionately affected by severe disease. Three of the 12 oximeter brands tested performed within acceptable bias limits across all Fitzpatrick groups; two did not. Regulators should require mandatory skin-tone-stratified accuracy data for all pulse oximeter devices seeking clinical clearance, and devices failing the Fitzpatrick 5-6 accuracy criterion should have their clinical indication restricted accordingly.

5.3 Limitations

Three limitations apply. First, device assignments were based on clinical indication rather than randomisation, and between-site variation in patient population characteristics may confound category-level accuracy comparisons. Sensitivity analyses restricted to matched subgroups showed consistent direction but slightly attenuated demographic equity findings. Second, the pulse oximeter validation used wrist-worn devices only; finger-probe oximeters -- which use a different optical path -- were not included and may show different skin-tone accuracy profiles. Third, the 2020-2021 recruitment period overlapped with COVID-19 pandemic conditions affecting routine clinical care at some sites, potentially influencing patient flow and clinical workflow integration

ratings.

6. Conclusion

6.1 Key Findings

Among five wearable device categories validated across 14,284 patients, multiparameter biosensor patches (WCVS 0.906) and single-lead ECG patches (WCVS 0.891) meet clinical deployment standards with diagnostic accuracy, equity, and workflow integration profiles sufficient to support clinical prescribing. CGM devices (WCVS 0.877) are clinically effective but require caveats for hypoglycaemia detection accuracy and anaemia. PPG smartwatches (WCVS 0.828) are screening tools, not diagnostic instruments, particularly in patients with ectopy. Wrist pulse oximeters (WCVS 0.762) carry clinically significant skin-tone-dependent bias -- up to +3.6 percentage points overestimation in dark-skinned patients -- that represents an unresolved patient safety concern requiring regulatory attention.

6.2 Recommendations

Three recommendations follow. First, clinical prescribing guidelines for wearable AF detection should distinguish between ECG patch devices (diagnostic quality, suitable for anticoagulation decision support) and PPG smartwatches (screening quality, requiring ECG confirmation before treatment decisions). Second, European and national regulatory bodies should mandate Fitzpatrick-stratified accuracy data for all pulse oximeter and PPG device submissions, with clear accuracy criteria for Fitzpatrick 5-6 skin tones as a condition of clinical clearance. Third, CGM prescribing guidance should explicitly note reduced hypoglycaemia detection accuracy in patients with haematocrit below 30% and in the first 24 hours of sensor wear, prompting supplemental blood glucose checks in these higher-risk circumstances. All validation data and the WCVS calculator are available from the corresponding author.

References

- Beck, R. W. et al. (2017). Effect of continuous glucose monitoring on glycemic control in adults with type 1 diabetes using insulin injections. *JAMA*, 317(4), 371-378.
- Bent, B. et al. (2021). The digital biomarker discovery pipeline: An open-source software platform for the development of digital biomarkers using mHealth and wearables data. *Journal of Clinical and Translational Science*, 5(1), e19.

Boldo, A. et al. (2022). Closed-loop insulin delivery improves glycaemic outcomes in type 1 diabetes. *Endocrinology and Diabetes*, 38(4), 412-421.

Dagher, L. et al. (2020). Wearable patch monitors in heart failure: A review. *JACC Heart Failure*, 8(6), 450-466.

Dolan, E. et al. (2021). Ambulatory blood pressure monitoring in clinical practice. *American Journal of Hypertension*, 34(7), 653-662.

Freedman, N. et al. (2020). Clinical applications of continuous glucose monitoring in non-insulin-treated patients. *Diabetes Technology and Therapeutics*, 22(S1), S15-S22.

Gladstone, D. J. et al. (2014). Atrial fibrillation in patients with cryptogenic stroke. *New England Journal of Medicine*, 370(26), 2467-2477.

Hannun, A. Y. et al. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25(1), 65-69.

Kaminska, M. et al. (2015). Wearable monitoring of vital signs in post-acute care. *Journal of Clinical Monitoring and Computing*, 29(5), 643-651.

Luzum, J. A. et al. (2021). Wearable heart rate monitors in clinical cardiology. *JACC Clinical Electrophysiology*, 7(5), 676-686.

Obermeyer, Z. et al. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.

Perez, M. V. et al. (2019). Large-scale assessment of a smartwatch to identify atrial fibrillation. *New England Journal of Medicine*, 381(20), 1909-1917.

Poole-Wilson, P. A. et al. (2003). Comparison of carvedilol and metoprolol on clinical outcomes in patients with chronic heart failure. *The Lancet*, 362(9377), 7-13.

Sjoding, M. W. et al. (2020). Racial bias in pulse oximetry measurement. *New England Journal of Medicine*, 383(25), 2477-2478.

Svensden, J. H. et al. (2021). Implantable loop recorder detection of atrial fibrillation to prevent stroke (LOOP). *The Lancet*, 398(10310), 1507-1516.

Turakhia, M. P. et al. (2019). Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study. *American Heart Journal*, 207, 66-75.

Ware, P. et al. (2020). Designing and evaluating a wearable-based remote monitoring program for heart failure. *Clinical Medicine Insights: Cardiology*, 14, 1179546820929471.

Wenger, N. K. et al. (2022). Cardiovascular health equity: Women and underrepresented minorities in cardiac trials. *JACC Advances*, 1(1), 100008.

Witkowski, M. et al. (2022). Gut microbiota and cardiovascular disease. *Circulation Research*, 130(5), 725-744.

Yilmaz, D. et al. (2021). Accuracy of consumer wearables for heart rate measurement during exercise. *Journal of Sports Sciences*, 39(12), 1398-1408.

Zimetbaum, P. and Goldman, A. (2010). Ambulatory arrhythmia monitoring: Choosing the right device. *Circulation*, 122(16), 1629-1636.

Declarations

Funding

Funded by the Spanish Ministry of Science and Innovation (PID2021-124631OB) and the Swedish Innovation Agency VINNOVA (2021-02846). No funder influenced study design, device selection, data analysis, or publication.

Conflict of Interest

All authors declare no competing interests and no financial relationships with any wearable device manufacturer or distributor evaluated in this study.

Data Availability Statement

Aggregated device-level accuracy and equity data are available from the corresponding author. Patient-level physiological data are subject to institutional data governance agreements. The WCVS calculation tool is publicly available on request.

Ethical Approval

Approved by Hospital La Paz Madrid Research Ethics Committee (PI-2019-584), Karolinska Ethical Review Authority (2019-05244), Charite Ethics Committee Berlin (EA1/284/19), Amsterdam UMC Ethics Board (NL71846.018.19), and Vienna General Hospital Ethics Committee (EK-2019-1284). Written informed consent obtained from all participants.

Appendix A

WCVS Scoring Rubric and Device Technical Specifications

Full WCVS dimension scoring criteria and technical specifications for each validated wearable device category.