

Wearables, Biosignals, and Remote Patient Monitoring

Study Guide

by Richard J. Young, Ph.D.

richard@deepneuro.ai

Chapter 14 (enrichment reading) of "Healthcare Analytics and AI: Building Systems That Actually Work"
The essentials: what to know, why it matters at the bedside, and how models fail.

One-sentence thesis. A wearable becomes clinically useful not when it detects a signal but when that signal can be turned into timely, trustworthy, equitable action.

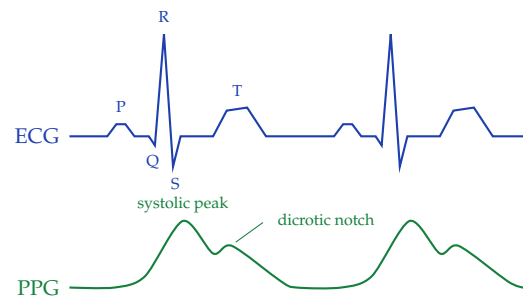
PART 1: Cardiac AI, PPG and ECG

Two Signals, Complementary Not Redundant

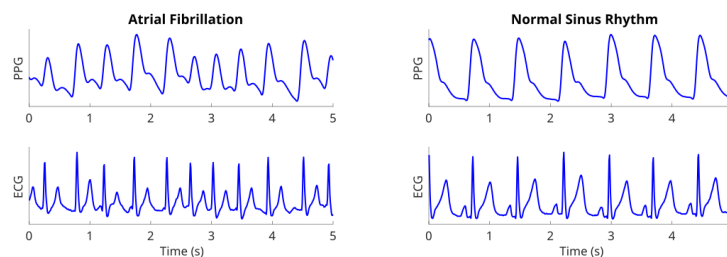
Reading a heart has always had a scheduling problem: the dangerous rhythm rarely shows up during the ten-minute clinic visit. Atrial fibrillation is often paroxysmal, coming and going, so an episodic 12-lead ECG can look perfectly normal an hour before a stroke. That is the structural advantage a worn sensor has over any clinic test, it is always on, and it is why wearables matter for cardiac care at all.

Wearables listen to the heart in two different languages. The ECG measures the heart's electrical activity through skin electrodes; the PPG (photoplethysmography) measures blood-volume changes by shining light into the skin and reading how much bounces back. Every smartwatch uses PPG for continuous heart rate; some add a single-lead ECG on the crown. The two are complementary, not redundant: ECG is

the wiring diagram of the heart's electrical system, while PPG watches the pipes swell and shrink with each pump. Fusing the electrical and hemodynamic features beats reading either signal alone.



Two complementary signals. ECG reads the heart's electrical wiring (the P-QRS-T complex); PPG reads the blood-volume pulse it drives (systolic peak, then diastolic notch). Fusing electrical and hemodynamic features beats either alone.



Atrial fibrillation vs normal sinus rhythm, in both PPG and ECG. The irregular rhythm (left) is what a continuously worn sensor catches and an episodic clinic ECG misses. Image: P. Charlton, CC BY 4.0.

AFib Detection: Reading the Numbers Honestly

The question a patient asks of a smartwatch is blunt: can it actually catch a dangerous rhythm? The Apple Watch earned FDA De Novo clearance for irregular-rhythm notification in 2018, and the pooled numbers are respectable, roughly 83 percent sensitivity and 79

percent specificity with the original PPG algorithm, improving to about 90 / 92 with updated versions that add the single-lead ECG. Those are legitimate screening numbers for a modifiable stroke risk factor.

ANALOGY

A metal detector at the beach: good enough to tell you where to dig, not good enough to tell you it is treasure. The confirmatory 12-lead ECG is the shovel.

The trap is reading more into the numbers than they carry: “cleared for AFib screening” is not “validated for comprehensive cardiac health.” The research fron-

tier is genuinely striking, attention-based models can translate a PPG trace into a proxy ECG (PR-AUC 0.986 versus 0.987 for a real ECG), and an AI-PPG estimate of vascular age flags cardiovascular risk when it runs ahead of a person’s chronological age (hazard ratio 2.37 once the gap reaches nine years or more), all with no blood draw. But a striking research result is not a cleared device.

Study	Signal(s)	Sensitivity	Specificity
Apple (original)	PPG	83%	79%
Apple (updated)	PPG + ECG	90%	92%
Fitbit Heart Study	PPG	67%	98%

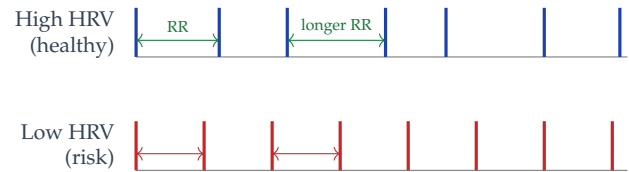
Heart Rate Variability

A healthy heart is not a metronome. Between beats it speeds up and slows down constantly, flexing with each breath like a jazz drummer bending tempo, and that flex is the point. Heart rate variability (HRV) is the beat-to-beat variation in the interbeat intervals, and it reflects the balance of sympathetic and parasympathetic tone. Counterintuitively, more variability is the healthy sign: higher HRV generally signals resilience, while reduced HRV tracks cardiovascular disease, diabetes, depression, and even mortality. When the flex disappears and the beats march like a metronome, the system is stiffer, not stronger.

The standard summary is RMSSD, the root mean square of successive differences between adjacent beats. In symbols:

$$\text{RMSSD} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (RR_{i+1} - RR_i)^2}$$

HRV is a genuinely dual-use biomarker, surfacing autonomic dysfunction in Long COVID and craving in addiction treatment, but consumer HRV is easily corrupted by motion artifacts, poor skin contact, and posture, all of which the preprocessing pipeline must handle before any feature can be trusted.



HRV as a rhythm strip. A healthy heart is a jazz drummer: the RR intervals flex (top). When that flex disappears and beats march like a metronome (bottom), the system is stiffer, not stronger. RMSSD quantifies the flex.

Sleep Staging from a Wristwatch

The bottleneck in sleep medicine is access, not science. Polysomnography costs \$1,500 to \$3,000 and carries a months-long waitlist, while 30 million Americans have obstructive sleep apnea and **80 percent** are undiagnosed. The opening is that heart rate and HRV shift systematically across sleep stages, so a PPG signal alone can approximate staging: SleepPPG-Net reached a Cohen's Kappa of 0.75 for four-class staging. That is like guessing which movie is playing next door from the rhythm of the music through the wall, not the full picture, but enough to route the right patients to the sleep lab. A wrist device that screens for apnea can triage the high-probability cases to an in-lab study, cutting wait times and reaching the roughly 24 million who are currently undiagnosed.

GluFormer and Foundation Models for Biosignals

The same shift that transformed language modeling is now reaching sensors: pretrain one big model on

oceans of unlabeled signal, then adapt it to many tasks. GluFormer (*Nature*, 2025) is a transformer foundation model trained self-supervised on 10 million glucose readings, transferring across 19 cohorts, 5 countries, and 8 devices. It out-predicted HbA1c for diabetes onset, capturing **66 percent** of new diagnoses in its top risk quartile versus 7 percent in the bottom. The value is that a single pretrained model behaves like one well-read intern who has seen every chart in the hospital and can be reassigned to any ward with a day of orientation, rather than training a new specialist per task. LSM-2 pushes this further: its Adaptive Inherited Masking learns directly from *incomplete* data, reading around gaps the way a clinician reads around missing labs instead of pretending they exist. **Watch out.** Model development is outrunning the clinical-validation pipeline of prospective trials, regulatory review, and workflow integration. Capability is not clearance.

Model	Scale	Modalities	Key innovation
LSM-2	40M hours	HR, HRV, accel, EDA, temp	Masking for missingness
SensorLM	59.7M hours	Sensor + text	Language alignment (zero-shot)
NormWear	Multi-source	PPG, ECG, EEG, GSR, IMU	Cross-signal pretraining

Subject-Level Leakage: The Trap That Fakes 98 Percent

The most seductive number in wearable ML is often a lie of the split. If 10-second segments from the same patient land in both the training and test sets, the model memorizes the *person*, not the disease, and a reported per-segment accuracy of **98 percent** can collapse the moment it meets a new face. It is like studying for an exam by memorizing the answer sheet with the students' names still on it: you ace the practice test and fail the real one. Worse, a model that scores brilliantly per segment may still fire dozens of false alerts per day per patient, because clinicians act on episodes, not windows. The fix is discipline: always split by subject (StratifiedGroupKFold on patient ID) and evaluate at the *episode* level, not the segment level.

Skin-Tone Bias Is Built Into the Hardware

Some wearable bias is not in the model at all, it is in the physics. Green-light PPG is absorbed more strongly by melanin, so some watches underestimate heart rate by 10 to 15 bpm at rest and by more than **20 percent** during vigorous activity in darker-skinned users, and most validation cohorts underrepresent Black participants. It is a camera whose auto-exposure was tuned only for light faces: not neutral, but failing specific people by design. The clinical stakes compound the injustice, an arrhythmia detector trained mostly on lighter skin will perform worst for the communities already facing the greatest health disparities. The mitigation is engineering, not a disclaimer: diverse training data and multi-wavelength sensors, not a warning label.

Reimbursement Is What Decides Deployment

A sensor that works but has no billing code is a sensor that stays in the lab; the CPT code, not the hardware, is the on-switch. CMS pays for remote patient monitoring (RPM) through codes covering setup, device supply, and clinician time, and the arithmetic is what funds the entire industry: roughly **\$110,000** in annual

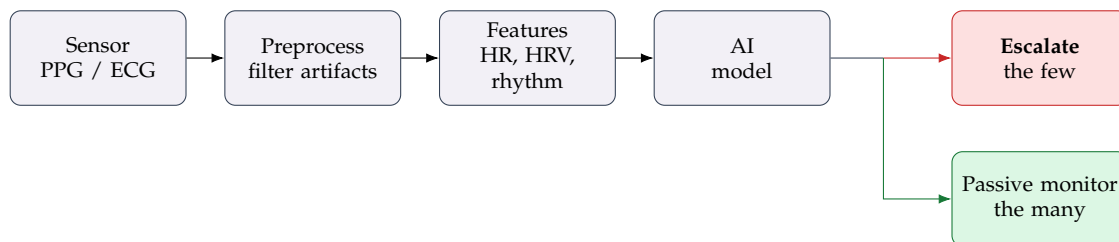
reimbursement per 100 enrolled patients. New 2026 codes (99445, 99470) extend RPM to shorter monitoring windows and to acute or post-surgical care. The clinical payoff is real, too, fusing wearable data with the EHR improves outcome prediction by **+8.5 percent** AUROC on average across ten outcomes in *All of Us*, precisely because wearables capture the life *between* the 15-minute visits.

CPT	Covers	Requirement
99454	Device supply + transmission	16+ days/month
99457	First 20 min clinician time	Per month
99445	Short-duration RPM (new 2026)	2–15 days/30

The Workflow Gap

The technical ability to monitor is mature; the organizational ability to *act* on the data is not. A program for 500 heart-failure patients generates a continuous stream that someone must triage, and installing 500 doorbells that all ring in one room does not help if no one can tell which ring matters, the guard sim-

ply stops answering. **Watch out.** Without AI-driven triage the home just replays the hospital's alert fatigue (Chapter 6). The sustainable pattern is continuous AI monitoring with *selective* human escalation: surface the few deteriorating patients, and let the model watch the rest passively. The signal becomes care only at that final human step.



The sustainable RPM pattern. Continuous AI monitoring with selective human escalation: flag the few deteriorating patients (red), monitor the rest passively (green). The signal becomes care only at the final human step.

Edge AI: Access, Not Optimization

Wearables run on tiny batteries, so streaming all raw data to the cloud is impractical, which pushes the model onto the device itself. TinyML runs inference locally (95 percent AFib accuracy with 7-day battery life), and on-device transcription with a 1-billion-parameter model keeps patient data off the network entirely. Frame this as access, not performance tuning: for the **21 percent** of rural Americans without reliable broadband, on-device inference is a pocket calculator that still works in a blackout, versus a website that needs the power grid up just to add two numbers. It is a prerequisite for reaching those patients at all, and it keeps sensitive data local by design (Chapter 3).

Who Benefits, Who Is at Risk

The same technology cuts both ways. Chronic-condition patients gain **30 to 50 percent** readmission reductions; rural and elderly populations gain reach and aging-in-place safety. But the risks are just as concrete: round-the-clock data privacy exposure, skin-tone bias, alert fatigue transplanted into the home, and the “worried well.”

▲ THE WORRIED-WELL TRAP

A wearable can be a smoke detector so sensitive it screams at steam from the shower, turning healthy people into anxious patients, orthosomnia, the anxious pursuit of a perfect sleep score, is now a recognized phenomenon.

The uncomfortable truth is that the same device which prevents a stroke through early AFib detection can degrade quality of life through over-monitoring, and a home flooded with daily alerts teaches the patient to ignore the one that finally matters.

Should You Build This?

Build it when the signal routes to an action someone will take inside the window where it still matters: an irregular-rhythm notification that reaches a clinician who can confirm and anticoagulate, a glucose trend that changes a dose, an RPM program with AI triage and a named person on the escalation path. The reimbursement pathway exists, and the readmission reductions in well-designed heart failure programs came from that closed loop, not from the sensor. **Think twice when** there is no triage layer, when you are extrapolating past a narrow clearance (cleared for AFib screening is not cleared for general health assessment), when green-light PPG goes to darker-skinned users without validating the known accuracy gap, or when the target is healthy people who will find anxiety in normal physiological variation. **Who bears the risk:** the patient absorbs both sides of the error, the false alarm that erodes trust and the real event lost in the noise; users whose skin tone was underrepresented in validation bear a miscalibration nobody measured; and the program bears the cost of data no one has time to read.

Quick Self-Test

1. Why do PPG and ECG features together outperform either alone for AFib detection?
2. A wearable AFib model reports 98 percent segment-level accuracy. What must you check before believing it, and why can it still be clinically useless?
3. Explain skin-tone bias in green-light PPG and name a mitigation that is not just a warning label.
4. Give the RMSSD idea in one sentence and say what low HRV suggests clinically.
5. What is the “workflow gap” in RPM, and what design pattern keeps RPM from replicating hospital alarm fatigue?
6. Why is edge AI a prerequisite for access rather than a performance optimization in rural care?

If You Remember One Thing

The wearable revolution is an integration problem, not a sensor problem. Sensors and models increasingly work; the hard, unfinished part is connecting a sleep score or glucose forecast to a reliable clinical action that is timely, equitable, and free of alert fatigue. The device generates the data. The surrounding system decides whether that data becomes care.

Richard J. Young, Ph.D. · richard@deepneuro.ai · Senior AI Research Scientist, UnitedHealth Group · UNLV Lee Business School · ORCID 0000-0002-1109-7552