Validity, Reliability, and Regulatory Considerations for Consumer Devices in Clinical Research Data Collection

Consumer-grade wearables offer promise for reducing trial burden and improving engagement, but accuracy and reliability vary substantially by device, measurement, and population, requiring rigorous fit-for-purpose validation and careful endpoint selection before integration into regulated research.

CLINICAL TRIAL DATA collection is changing substantially. Consumer-grade wearable devices, including smartwatches, fitness trackers, and smart rings from manufacturers such as Apple, Fitbit, Garmin, and Oura, have grown rapidly. Over 4,000 clinical studies registered globally now use wearable technology as of 2023.[1] These tools offer promising avenues for capturing objective, longitudinal outcomes measures with minimal participant burden and may improve engagement, streamline data collection, and reduce the need for frequent site visits. However, integrating consumer devices into regulated clinical research environments presents important methodological, technical, and regulatory challenges.

This article explores key considerations for collecting data using consumer-grade devices in clinical trials, including their accuracy and reliability, comparability to data collected by research-grade devices, data security and privacy, and regulatory acceptance.

ABOUT THE AUTHORS

LAUREN CROOKS, MSC, Director, Science & Medicine, Signant Health; **GRETA MARIE VAN SCHOOR, PHD,** Associate Director, Science & Medicine, Signant Health; **ANTHONY EVERHART, MD,** Clinical Vice President, Internal Medicine, Signant Health; **BILL BYROM, PHD,** Vice President, Product Intelligence & Positioning, Signant Health; and **THIJS SONDAG,** Head of Product, WeGuide

Measurement accuracy and performance

Consumer devices are largely designed for general wellness, fitness, and lifestyle tracking rather than for clinical measurement to support regulatory decision making. Many therefore do not undergo the same strict validation and regulatory review process required for research-grade devices or devices with medical device classification.[2] As a result, performance can vary considerably, impacting both the validity and reliability of the data generated. Sources of potential variability include:[7]

- **Sensor technology and model differences:** Sensor type, design, sampling frequency, and embedded accuracy vary across devices. For example, heart rate may be monitored using photoplethysmography (PPG) or electrical bioimpedance (EBI), leading to manufacturer-specific accuracy profiles.[3]
- **Software and firmware upgrades:** Mid-study updates to device algorithms, calibration methods, or data processing routines can alter outputs, and affect consistency.[4,5]
- **Device placement and user behavior:** Inconsistencies in how and where devices are worn can significantly influence measurements such as heart rate, physical activity, or sleep patterns.[4,6]
- **Data transfer and handling:** Multi-step syncing from device to app to trial database can introduce missing data or complete data loss due to failure or delay in data syncing or interruptions in connectivity.[4,5]
- **Data access and transparency:** Restricted raw data access and "black box" algorithms limit independent verification, reproducibility, and validation.[20,21]
- **Proprietary measures:** Consumer wearables often measure similar underlying physiology but package it into different, non-equivalent proprietary scores, making direct comparison difficult.[7,21]
- **User physiology:** Factors such as skin tone, movement types, or activity levels can influence sensor performance and data accuracy.[6,13]

Some physiological parameters, such as heart rate, have demonstrated high concordance with clinical standards, whereas others, such as respiratory rate, skin temperature, and sleep staging exhibit considerable variability.[7,12] For example, Apple Watch heart rate data has demonstrated near-concordance with ECG-based clinical standards, while gait metrics (e.g., step length, cadence) often show systematic calibration biases.[7,12] Sleep trackers also vary, where research-grade EEG-based headbands, such as the Dreem smart headband, outperform other wearables in terms of staging accuracy, and consumer devices, such as Oura and Fitbit, show high variability in sleep stage tracking performance.[14,15]

To limit accuracy concerns when implementing consumer devices in clinical trials, best practice recommendations include:[10,11,12]

- Selecting fit-for-purpose devices that have demonstrated evidence of measuring the concept of interest in similar populations to the trial population.[10]
- Locking device firmware and hardware versions across study duration.[11]
- Provide standardized use instructions and run-in adaptation periods.[11]
- Select clinically meaningful endpoints first to drive the choice of device with proven validity for that outcome.[10,22]

- Employ ongoing monitoring of performance drifts and data quality throughout the trial duration.[11,22]

Ultimately, the acceptability of data collected using consumer devices depends on the intended use within the trial. Data used for exploratory endpoints or monitoring adherence may not require the same level of accuracy as data supporting the primary efficacy or safety outcomes. A rigorous, fit-for-purpose assessment aligned with regulatory guidance should guide decisions around device selection and endpoint development.[10,11,22]

Reliability and consistency

Beyond point accuracy, reliability and consistency of measurements across time, conditions, and populations are critical for longitudinal clinical trials.[7,18]

Movement artifacts pose the primary challenge for PPG-based measurements. While consumer devices employ machine learning algorithms to filter motion-related noise, accuracy degrades during physical activity. Mean absolute error for heart rate increases from under 2 bpm during rest to 5 to 8 bpm during moderate activity and over 10 bpm during vigorous exercise.[12,13] Device-to-device variability within the same manufacturer and model introduces additional uncertainty, with differences in sensor calibration, firmware versions, and proprietary algorithm contributing to variation in energy expenditure and step counts.[4,5]

User characteristics meaningfully influence measurement reliability. Body mass index (BMI), skin tone, age, and wrist circumference affect PPG signal quality and measurement accuracy. However, recent multicenter validation studies stratifying results across demographic variables found that modern consumer devices maintain acceptable accuracy (ARMS within predefined thresholds) across diverse populations. Performance remains slightly degraded in individuals with higher BMI or darker skin tones.[6,13]

Environmental factors such as ambient temperature, altitude, and humidity can also influence sensor performance, particularly for devices measuring skin temperature and oxygen saturation.[9,12]

Data comparability and systematic biases

While point estimates of accuracy are useful, understanding systematic biases and their implications for detecting treatment effects is critical in clinical trials.[7,18]

GROUP LEVEL VERSUS INDIVIDUAL LEVEL AGREEMENT

A consistent finding across validation studies is the discrepancy between group level and individual level measurement agreement. Consumer devices often show acceptable accuracy when data are aggregated across populations but exhibit substantial individual variability. For example, while the Oura Ring showed mean total sleep time (TST) differences of less than 12 minutes at the group level compared to PSG, individual level differences frequently exceeded ±60 minutes.[14]

This has direct implications for clinical trials. Group level endpoints may be adequately captured by consumer devices. However, individual patient responses, which are essential for personalized medicine approaches and safety monitoring, may be obscured by measurement variability.[14,18]

SYSTEMATIC BIAS PATTERNS

Consumer wearables may exhibit non-random, systematic biases that cannot be easily corrected through simple linear adjustments. Sleep tracking devices often overestimate sleep and underestimate wakefulness, while activity trackers may underestimate low intensity movement and overestimate high intensity activity.[15,18]

IMPACT ON TREATMENT EFFECT DETECTION

The practical consequence of measurement bias is illustrated by a validation study of the Garmin vívosmart® HR in a randomized trial of suvorexant for insomnia in Alzheimer's disease patients. While PSG detected a 35-minute increase in TST ($p = 0.057$), the consumer wearable underestimated this effect, detecting only a 20-minute difference ($p = 0.405$).[16] This attenuation of treatment effects due to measurement error can result in false negative trial results and failed development of effective therapies.

By contrast, in the RATE-AF trial comparing digoxin versus beta-blockers for heart rate control in atrial fibrillation, consumer wearables successfully captured over 143 million heart rate measurements. They showed equivalent treatment effects to intermittent ECG monitoring, validating their utility for this specific application.[17]

VALIDATION STANDARDS

A comprehensive living systematic review found that about only 11% of commercially available wearable devices have been validated for at least one objective physiological outcome against accepted reference standards. More critically, because typical devices measure multiple outcomes, the validation studies conducted to date represent only 3.5% of the total needed for comprehensive evaluation of device-outcome combinations.[7]

The lack of standardized validation protocols further complicates comparisons across devices and studies. Validation methodologies vary widely in terms of reference standards, statistical approaches (Bland-Altman vs. correlation coefficients vs. equivalence testing), population characteristics, and environmental conditions, limiting the ability to identify the most accurate devices for specific contexts of use.[7,18]

Data Privacy and protection of consumer- and research grade devices

Data security, privacy, syncing, export, informed consent, user rights, and data ownership are all important considerations when integrating consumer- and research-grade devices into clinical trials.[18,19,20] The use of wearables introduces technical, legal, and ethical challenges.[20,21]

Clinical trials that incorporate wearable technology are subject to stringent data privacy regulations and standards. In the European Union, the General Data Protection Regulation (GDPR) governs how wearable device data should be processed and requires robust security measures, clearly defined purposes for data use, consent for data reuse and sharing, and have specific requirements around consent.[18,20] It also grants participants specific rights, including access to their data, as well as the ability to rectify or erase personal information.

Similarly, federal laws such as Health Insurance Portability and Accountability Act (HIPAA) in the US regulates how such personal health information should be collected, stored, used, and shared, when the wearable device participant data is collected or managed by covered entities. HIPAA requires patient consent for data collection and sharing, along with the implementation of technical and administrative safeguards such as encryption and access controls.

Data collected from both research-grade and consumer-grade devices in clinical trials is subject to oversight by study sponsors, Institutional Review Boards, and regulatory authorities. The FDA's *Digital Health Technologies for Remote Data Acquisition in Clinical Investigations* guidance outlines recommendations for data collection methods and maintenance of data integrity and security.[22] Similarly, the EMA's *Guideline on Computerized Systems and Electronic Data in Clinical Trials* emphasizes data integrity and confidentiality.[23] When a wearable is classified as a medical device, additional regulatory considerations apply. Sponsors must ensure the device has undergone appropriate validation and regulatory clearance, demonstrating technical performance, clinical relevance, and reliability. The level of performance acceptable for a consumer device to obtain market approval may be lower than what regulators expect for devices used as research tools in clinical trials, where higher accuracy and precision are often required to support regulatory decision making.

Consumer-grade devices often require additional privacy and security adaptations because personal identifying information is commonly collected.[24] Users may need to share names or email addresses when enrolling, and certain device data, such as GPS location, can be inherently identifiable. Commercial wearable devices can also include features such as location tracking or data sharing via social media that should be disabled to protect participant privacy and ensure safe collection and use of data. When using these devices in clinical trials, collecting specific personal information is often prohibited, so the necessary adaptations should be made to meet applicable privacy requirements.[24]

Platform solutions such as WeGuide address many of these privacy and compliance challenges through direct device connectivity. Rather than routing data through manufacturer cloud servers, WeGuide connects to wearable devices such as Garmin directly via Bluetooth on the participant's device. This approach allows sensitive health data to flow directly from the wearable to the clinical trial database without intermediate storage on commercial platforms, helping sponsors maintain greater control over data handling and simplifying compliance with regulations such as GDPR and HIPAA. This direct connection model also eliminates concerns about manufacturer server outages or policy changes that could disrupt data collection mid trial.[25]

Study-specific consent language may also be required. In many cases, these devices lack default mechanisms for raw data access or advanced security, which necessitates additional safeguards, encryption, and secure transfer mechanisms to protect participants' sensitive health data.

By contrast, research-grade devices are designed with built-in protections that align with clinical and legal standards. They typically provide full access to raw data, along with integrated security features compliant with stringent regulations. These devices support transparent data management with encryption, user authentication, consent tracking, and audit trails, enabling ethical data handling and regulatory compliance throughout the research process.

Guidance from the US National Institute of Standards and Technology (NIST) further strengthens best practices for data privacy and security, to safeguard personal data.[20] Following NIST guidance helps to ensure robust encryption, secure device management, data traceability, risk-based controls, and clear patient consent processes. Adhering to these standards helps mitigate legal and ethical risks while ensuring responsible handling of participant data in trials.

In summary, there are some differences around data privacy and protection between consumer devices and research-grade devices, associated with regulatory requirements, data accuracy, and management protocols. Research-grade devices offer more robust protections and greater reliability. However, both device types must ensure participant confidentiality, data integrity, and compliance with privacy regulations. Importantly, proper privacy and data protection protocols are essential in clinical trials regardless of device type.

Global regulatory framework

As the capabilities of wearables expand, regulatory frameworks are evolving in parallel to safeguard data integrity, participant safety, and privacy—anchored by the standards and guidelines of the FDA and the EMA.[22,23]

FDA

At the heart of the FDA's *Digital Health Technologies for Remote Data Acquisition in Clinical Investigations* guidance is a risk-based approach coupled with fit-for-purpose device validation. Digital health technologies used as clinical trial tools must demonstrate reliability, accuracy, and suitability for measuring clinically meaningful endpoints. Studies employing wearables that inform treatment decisions or capture safety or efficacy data draw heightened regulatory scrutiny, with device classification determining the approval pathway and oversight intensity. The FDA also mandates robust data management, encryption, and privacy protection, with informed consent and participant safety protocols embedded in trial design. The agency's *General Wellness: Policy for Low Risk Devices* does not exempt products intended to measure or report physiologic values for medical or clinical purposes.[2,22]

EMA

The EMA's Guideline on Computerized Systems and Electronic Data in Clinical Trials emphasizes Good Clinical Practice compliance, including validation, reliability, and integrity of computerized systems and wearables used in clinical research. EMA expectations include system validation, data accuracy, clear audit trails, harmonized access controls, privacy safeguards, and GDPR-compliant data management, particularly in multinational studies with cross-border data transfers.[23]

DATA QUALITY

Both FDA and EMA stress the vital importance of data quality and the authenticity of real-world evidence generated by wearables. Sponsors must be vigilant about device calibration, data normalization, and error-detection protocols in the face of high-frequency, high-volume data streams. Advanced quality assurance, from real-time validation checks to cross-device harmonization, ensures the reliability and regulatory acceptability of wearable-derived datasets.[22,23]

DATA PRIVACY

Privacy and participant rights anchor every phase of wearable device use in clinical trials. Informed consent, transparency regarding continuous monitoring, and ethics committee review are required to build participant trust and protect data sensitivity. Regulatory guidance mandates encrypted transmission, explicit consent, restricted access, and robust communication regarding device capabilities and limitations.[22,23]

PATIENT SAFETY

Patient safety is a non-negotiable regulatory priority. Both agencies require sponsors to anticipate and mitigate risks such as device malfunction, skin irritation, or non-compliance due to poor usability or training gaps. Regulatory considerations highlight the importance of maintaining high adherence rates by providing clear instructions, intuitive interfaces, and responsive patient support.[22,23]

Conclusion

As clinical trials continue to embrace decentralization and real-world data, wearables will remain pivotal in expanding trial reach, increasing patient engagement, and driving the evolution of regulatory frameworks. Ultimately, successful incorporation of sensors and wearables into clinical research demands rigorous device validation, data security, ethical oversight, and harmonized compliance with FDA and EMA guidance, safeguarding participants while maintaining the scientific integrity of trial results.

By integrating this fit-for-purpose validation, strong quality assurance, and comprehensive privacy and safety safeguards, clinical trial teams can responsibly harness both consumer- and research-grade wearables to generate reliable evidence and support evolving FDA and EMA expectations for digital health technologies.

References

1. Lampreia, F., Madeira, C., & Dores, H. (2024). Digital health technologies and artificial intelligence in cardiovascular clinical trials: A landscape of the European space. Digital Health, 10, 20552076241277703.
2. Food and Drug Administration. (2016, July 29). General wellness: Policy for low risk devices. Guidance for industry and Food and Drug Administration staff. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/general-wellness-policy-low-risk-devices
3. Lapsa, D., Janeliukstis, R., Metshein, M., & Selavo, L. (2024). PPG and bioimpedance-based wearable applications in heart rate monitoring—a comprehensive review. *Applied Sciences*, 14(17), 7451.
4. Van Der Donckt, J., Vandenbussche, N., Van Der Donckt, J., et al. (2024). Mitigating data quality challenges in ambulatory wrist-worn wearable monitoring through analytical and practical approaches. Scientific Reports, 14, 17545.
5. Böttcher, S., et al. (2022). Data quality evaluation in wearable monitoring. *Scientific Reports*, 12, 21412.
6. Bent, B., Goldstein, B. A., Kibbe, W. A., et al. (2020). Investigating sources of inaccuracy in wearable optical heart rate sensors. *npj Digital Medicine*, 3, 18.
7. Doherty, C., Baldwin, M., Keogh, A., Caulfield, B., & Argent, R. (2024). Keeping pace with wearables: A living umbrella review of systematic reviews evaluating the accuracy of consumer wearable technologies in health measurement. *Sports Medicine*, 54(11), 2907–2926.
8. Izmailova, E. S., McLean, I. L., Bhatia, G., Hather, G., Cantor, M., Merberg, D., Perakslis, E. D., Benko, C., & Wagner, J. A. (2019). Evaluation of wearable digital devices in a Phase I clinical trial. *Clinical and Translational Science*, 12, 247–256.
9. Mendt, S., Zout, G., Rabuffetti, M., Gunga, H. C., Bunker, A., Barteit, S., & Maggioni, M. A. (2025). Laboratory comparison of consumer-grade and research-established wearables for monitoring heart rate, body temperature, and physical activity in sub-Saharan Africa. *Frontiers in Physiology*, 16, 1491401.
10. Byrom, B., Watson, C., Doll, H., Coons, S. J., Eremenco, S., Ballinger, R., McCarthy, M., Crescioni, M., O'Donohoe, P., & Howry, C. (2018). Selection of and evidentiary considerations for wearable devices and their measurements for use in regulatory decision making: Recommendations from the ePRO Consortium. *Value in Health*, 21(6), 631–639.

11. Clinical Trials Transformation Initiative (CTTI). (2021). CTTI considerations for advancing the use of digital technologies for data capture & improved clinical trials. https://ctti-clinicaltrials.org/wp- content/uploads/2021/06/CTTI_Digital_Health_Technologies_Recs.pdf
12. Chan, P., Ryan, N., Chen, D., McNeil, J., & Hopper, I. (2022). Novel wearable and contactless heart rate, respiratory rate, and oxygen saturation monitoring devices: A systematic review and meta-analysis. *Anaesthesia*, 78(1), 1292–1306.
13. Chen, W., Cordero, R., Lever Taylor, J., Pangallo, D. R., Picard, R. W., Cruz, M., & Regalia, G. (2024). Multicenter evaluation of machine-learning continuous pulse rate algorithm on wrist-worn device. *Digital Biomarkers*, 8(1), 288–301.
14. Herberger, S., Aurnhammer, C., Bauerfeind, S., Bothe, T., Penzel, T., & Fietze, I. (2025). Performance of wearable finger ring trackers for diagnostic sleep measurement in the clinical context. *Scientific Reports*, 15, 5641.
15. Feehan, L., Geldman, J., Sayre, E. C., Park, C., Ezzat, A., Yoo, J. Y., Hamilton, C. B., & Li, L. C. (2018). Accuracy of Fitbit devices: Systematic review and narrative syntheses of quantitative data. *JMIR mHealth and uHealth*, 6(8), e10527.
16. Svetnik, V., Wang, T., Ceesay, P., Snyder, E., Ceren, O., Bliwise, D., Budd, K., Hutzelmann, J., Stevens, J., Lines, C., Michelson, D., & Herring, W. (2021). Pilot evaluation of a consumer wearable device to assess sleep in a clinical polysomnography trial of suvorexant for treating insomnia in patients with Alzheimer's disease. *Journal of Sleep Research*, 30(6), e13328.
17. Gill, S., Barsky, A., Guan, X., Bunting, K., Karwath, A., Țica, O., Stanbury, M., Haynes, S., Folarin, A., Dobson, R. J. B., Kurps, J., Asselbergs, F., Grobbee, D., Camm, A., Eijkemans, M., Gkoutos, G., & Kotecha, D. (2024). Consumer wearable devices for evaluation of heart rate control using digoxin versus beta-blockers: The RATE-AF randomized trial. *Nature Medicine*, 30, 2127–2135.
18. Huhn, S., Axt, M., Gunga, H. C., Maggioni, M. A., Munga, S., Obor, D., Sié, A., Boudo, V., Bunker, A., Sauerborn, R., & Bärnighausen, T. (2022). The impact of wearable technologies in health research: Scoping review. JMIR mHealth and uHealth, 10(1), e34384.
19. Doherty, C., Baldwin, M., Lambe, R., Altini, M., & Caulfield, B. (2025). Privacy in consumer wearable technologies: A living systematic analysis of data policies across leading manufacturers. *npj Digital Medicine*, 8(1), 363.
20. Izmailova, E. S., Wagner, J. A., & Perakslis, E. D. (2018). Wearable devices in clinical trials: Hype and hypothesis. *Clinical Pharmacology & Therapeutics*, 104(1), 42–52.
21. Canali, S., Schiaffonati, V., & Aliverti, A. (2022). Challenges and recommendations for wearable devices in digital health: Data quality, interoperability, health equity, fairness. *PLOS Digital Health*, 1(10), e0000104.
22. Final FDA guidance. (2023). Digital Health Technologies for Remote Data Acquisition in Clinical Investigations.
23. EMA guideline. (2023). Guideline on Computerised Systems and Electronic Data in Clinical Trials.
24. Rudo, T., & Dekie, L. (2024). The future fit of wearables for patient-centric clinical trials. *Applied Clinical Trials*.
25. WeGuide. (2025). Wearable device studies. https://www.weguide.health/wearable-device-studies