

Agreement and Interchangeability of Nightly Temperature Deviation Metrics Between Oura Ring and Garmin Wearable Devices

Implications for Illness Detection and Physiological Stress Monitoring

Calum Garrigan, MS³, Stephen Suydam, PhD³, Nathan Weiland, MS³, Carrie Pockrandt, DHSc², Ryan Carroll, MS³, Jonathan Zeisset, MS¹, Edward Segovia, BS², Paul Lane, MS³, Natasha Kreisle, BA¹, Aaron Anderson, BA², Brandon Marcello, PhD³, Jacqueline Ortiz MS, MMCI¹, Cory Bunnetto, BBA², Todd Norell, BA³, Madhu a Shrestha, MPH¹, Stephen DeLellis, MPAS², A.J. Reuter, BA³, Dennis McGurk, ^{1A} PhD¹, Julie Cantelon, PhD⁴, and Joseph Patterson⁴

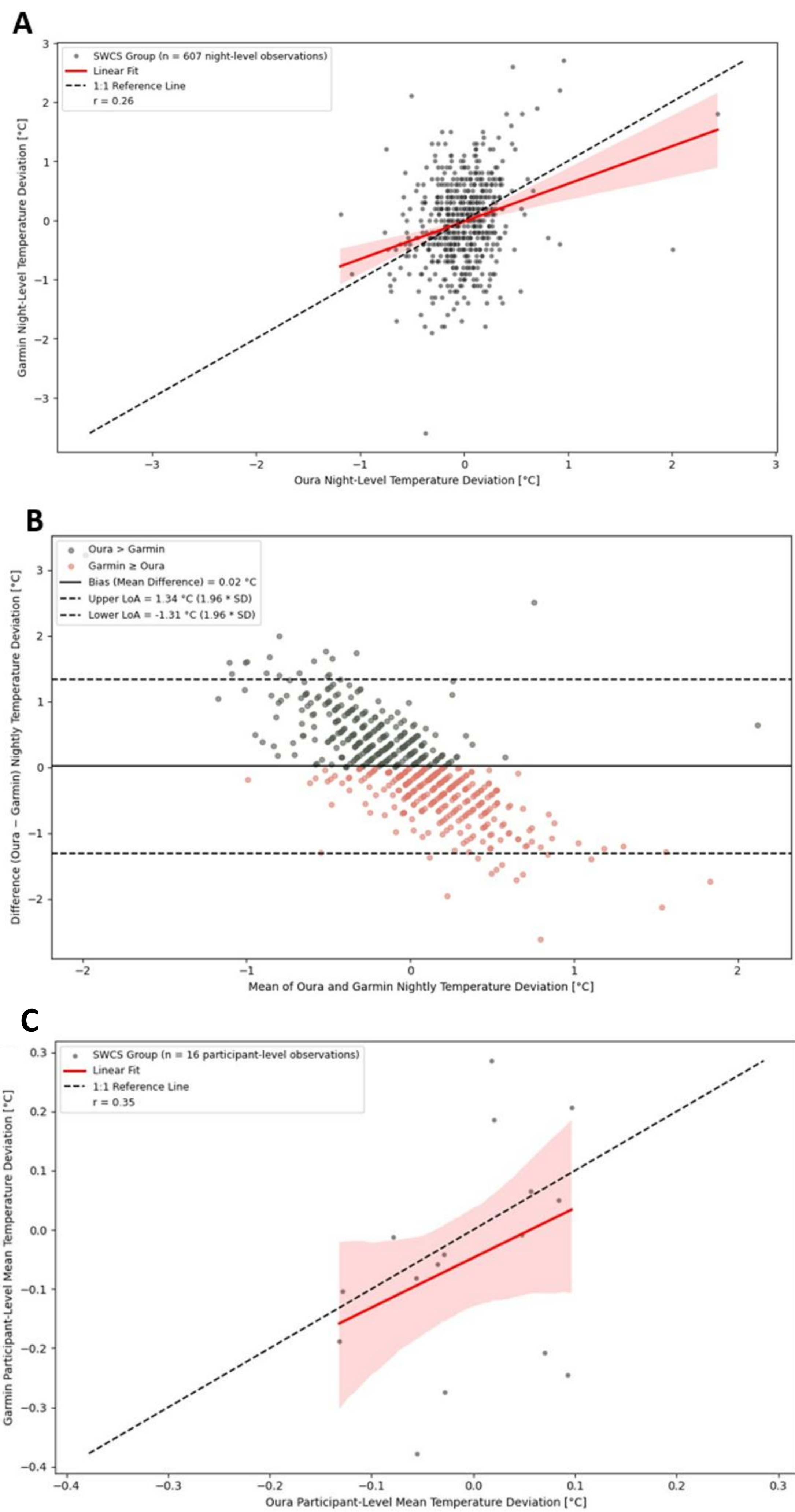


Figure 1. (A) Correlation plot of nightly temperature deviation values between Oura and Garmin devices ($r = 0.26$), with linear fit and 1:1 reference line shown. (B) Bland–Altman plot demonstrating a mean bias of 0.02 °C and limits of agreement of –1.31 °C to 1.34 °C. (C) Correlation plot of participant-level mean temperature deviation values between Oura and Garmin devices ($r = 0.35$), with linear fit and 1:1 reference line shown.

METHODS

Participants:

- N = 16 Service Members
- OHWS qualification course participants

Devices:

- Oura Ring Gen 3 and Gen 4 (Oulu, Finland)
- Multiple Garmin Watch Models (Garmin Ltd., Olathe, KS, USA)

Data Collection:

- Devices worn nightly; data synced via mobile apps
- Collected and managed via Teamworks Athlete Management System (AMS)
- Data processed to:
 - Remove duplicates
 - Consolidated repeated nightly records
- Final dataset: 607 paired nightly observations

Metrics Analyzed:

- Nightly temperature deviation (baseline-adjusted)

Timeframe:

- February 2025 – January 2026

Analysis Tools:

- Python (PyCharm, JetBrains)

Statistical Analysis:

- Pearson correlation (r)
- Bland–Altman plots: mean bias and Limits of Agreement (LoA = mean $\pm$ 1.96 SD)

CONCLUSION

This study found weak agreement between Oura Ring and Garmin wearable devices for nightly temperature deviation metrics. While mean bias was minimal, wide limits of agreement indicate substantial variability at the nightly level, preventing interchangeability for short-term monitoring.

Participant-level averaging improved agreement, suggesting both devices may capture similar long-term physiological trends. However, differences in device algorithms and sensor placement likely contribute to observed variability.

For operational use, these findings suggest that temperature changes should be interpreted using the same device over time, rather than comparing measurements between different devices. Future research can explore whether simple temperature thresholds could be used to help detect illness.

- Wearable devices are increasingly used to monitor physiological readiness in military populations
- The OHWS program utilizes wearables to monitor recovery, stress, and performance
- Nightly temperature deviation metrics are used to identify illness and physiological strain
- Devices report baseline-adjusted deviations, not absolute temperature values
- Wearables represent a significant operational cost, making cross-device compatibility important

THE AIM OF THIS STUDY is to evaluate agreement between Oura Ring and Garmin wearable devices for nightly temperature deviation metrics and determine whether these measurements can be used interchangeably.

OBJECTIVE

Describe how wearable devices report baseline-adjusted temperature deviation, representing device-generated deviations rather than absolute temperature measurements.

Analyze agreement between wearable platforms using correlation and Bland–Altman methods to determine whether temperature metrics are interchangeable.

Discuss implications of wearable temperature variability for illness monitoring and readiness assessment within military operational environments.

RESULTS

Night-Level Analysis

Comparison of paired Oura and Garmin temperature deviation values for each individual night across all participants.

- Correlation: **$r = 0.26$** → Weak agreement
- Mean Bias: **0.02 °C**
- LoA: **–1.31 °C to 1.34 °C**
- Indicates substantial night-to-night variability between devices

Participant-Level Analysis

Comparison of each participant's average Oura and Garmin temperature deviation across all available nights.

- Correlation: **$r = 0.35$** → Modest agreement
- Mean Bias: **0.05 °C**
- LoA: **–0.28 °C to 0.37 °C**
- Indicates reduced variability and improved agreement after averaging across nights

Observations

- Minimal systematic bias between devices
- Wide variability at the nightly level
- Improved agreement at the participant level
- Devices may reflect similar long-term trends
- Variability remains too large to support interchangeability