

Original Paper

Benefit of the N-of-1 Approach Versus Aggregate Analysis in Tracking Individual Trajectories During Pregnancy: Comparison of Longitudinal Wearable Observational Studies

Tina Behrouzi^{1,2,3*}, MSc; Jennifer Yu^{1,2,3,4*}, MSc; Robin Yang⁵, MSc; Adrien Boch^{6,7}, MSc; Anna Goldenberg^{1,2,3,8,9}, PhD; Sarah M Goodday^{5,10}, PhD; Stephen H Friend^{5,10}, MD, PhD

¹Department of Computer Sciences, University of Toronto, Toronto, ON, Canada

²Vector Institute, Toronto, ON, Canada

³Genetics and Genome Biology Program, Hospital for Sick Children, Toronto, ON, Canada

⁴Tri-institutional Computational Biology & Medicine, Weill Cornell Medicine, New York, NY, United States

⁵YouandMe, Great Neck, NY, United States

⁶Meta (United States), San Francisco, CA, United States

⁷Evidation Health (United States), San Mateo, CA, United States

⁸Department of Laboratory Pathology and Medicine, University of Toronto, Toronto, ON, Canada

⁹Canadian Institute for Advanced Research, Toronto, ON, Canada

¹⁰Department of Psychiatry, University of Oxford, Oxford, United Kingdom

*these authors contributed equally

Corresponding Author:

Stephen H Friend, MD, PhD

4YouandMe

185 Great Neck Road

Great Neck, NY 11021-3326

United States

Phone: 1 215-421-3073

Email: friend@4youandme.org

Abstract

Background: Personal digital health technologies (DHTs) enable real-time monitoring of physiological metrics and behavioral data, including heart rate variability (HRV), supporting analysis of pregnancy-related conditions and personalized care throughout the perinatal period. While recent studies demonstrate the utility of personal DHTs in tracking pregnancy-related symptoms, they often rely on aggregate statistical methods that overlook individual variability.

Objective: This study aims to compare aggregate and individual-level analyses of DHT data for pregnancy-related conditions, using the comprehensive BUMP (Better Understanding the Metamorphosis of Pregnancy) dataset to highlight the importance of individual variability and data heterogeneity.

Methods: We analyzed wearable and self-reported data from 256 participants enrolled in the BUMP study (January 2021 to May 2022), including HRV, sleep, and fatigue measured via Oura Rings and smartphone surveys. Individual-level (N-of-1) trajectories were evaluated and compared with aggregate results to uncover personal and collective trends. A statistical method was developed to assess the influence of adverse events and severe symptoms, while case studies explored confounding and modifying factors underlying heterogeneity. Comprehensive statistical analysis included the coefficient of determination, Kolmogorov-Smirnov tests, likelihood ratio tests, and Welch *t* tests, with interindividual variability flagged based on high-variability thresholds.

Results: Substantial interindividual variability was observed across all features. Only 4.76% (12/256) of participants exhibited an HRV inflection at the aggregate week-33 inflection point, with a coefficient of variation of 14.24%. The median value of the gestational week in individual fatigue troughs was 23 (IQR 8; range 8-38) weeks, differing from aggregate estimates. Distributional comparisons showed no statistically significant differences in individual-level model fit (R^2) by pregnancy complications or age (*P* values ranging from .06 to .99 across all model fit comparisons). Case studies further highlighted both intraindividual and interindividual differences, emphasizing the importance of considering external factors, such as adverse events and severe symptoms.

Conclusions: Our findings show that aggregate wearable data often fail to generalize across populations, oversimplifying pregnancy-related physiological and subjective changes. This simplification can obscure individual trajectories, leading to generalized insights that may not reflect many pregnant women's experiences. Our results highlight the impact of heterogeneity on pregnancy outcomes, emphasizing the need to move beyond one-size-fits-all models and leverage DHT for personalized care.

JMIR Form Res 2026;10:e86203; doi: [10.2196/86203](https://doi.org/10.2196/86203)

Keywords: digital health technologies and wearables; N-of-1 analysis and personalized insights; pregnancy; maternal health; advanced statistical modeling in health care

Introduction

Pregnancy is a critical period marked by significant physiological and psychological changes, influencing maternal and fetal health [1]. Conditions such as gestational hypertension, preeclampsia, preterm birth, and postpartum depression impose significant global health burdens [2]. Despite their prevalence, there is still a limited understanding of the underlying factors driving pregnancy outcomes and an essential need for improved support systems to address women's health during this critical transition [3].

Wearable devices, digital health apps, and other remote smart devices (personal digital health technologies [DHTs] [4]) are increasingly recognized for their potential to transform health care [5,6]. DHTs enable the semicontinuous, real-time monitoring of physiological measures and behavior information, including cardiovascular metrics, activity levels, sleep patterns, and high-frequency subjective symptoms [7-9], particularly relevant to pregnancy. These approaches offer an advantage over traditional aperiodic, symptom-monitoring methods in capturing subtle individual differences in a remote setting, filling in the gaps between clinic visits, and providing tools that could track key symptoms in populations who lack access to prenatal care [10,11]. Harnessing data from diverse DHTs could enhance early detection of pregnancy-related conditions [4,12,13] and enable personalized care throughout the perinatal period [14].

While recent studies have demonstrated the value of personal DHTs in tracking pregnancy-related symptoms and outcomes [15], they often rely on aggregate statistical methods [14,16-18] that can overlook individual variability, as a one-size-fits-all approach may not apply effectively to every individual [19]. A recent large-scale longitudinal study [20] demonstrates systematic deviations of sleep metrics from individual prepregnancy baselines across pregnancy and postpartum, providing population-level context in the field. However, variability in individual behaviors and physiological responses, along with limitations such as small sample sizes in some studies, continues to challenge the generalizability of population-based models. Personal DHTs in pregnancy offer individualized risk monitoring, but recent studies rely on aggregated data instead of N-of-1 analyses. Closing this gap is key to making N-of-1 a standard and maximizing wearable data in future research.

This paper aims to compare the generalizability of personal DHT study results derived from aggregate data with those based on N-of-1 analyses in pregnant individuals

using data from a US-based digital health pregnancy study (the BUMP [Better Understanding the Metamorphosis of Pregnancy] study [21]). Three recently published pregnancy digital health studies [16-18] were selected that focused on 3 key measures: heart rate variability (HRV) [16], sleep [18], and fatigue [17]. Aggregated results from these studies were replicated using the BUMP study dataset and were extended by exploring N-of-1 analyses of these key measures. By applying N-of-1 methods, including spline fitting, pregnancy condition analysis, and analyzing certain women in case studies, we aimed to dissect individual differences often overlooked in aggregated approaches.

Methods

Study Design

The BUMP study was a participant-centric digital health study that tracked maternal symptoms using wearable devices and smartphone apps from preconception, through pregnancy, and up to 3 months postpartum [21]. The study was conducted fully remotely and designed to capture high-frequency physiological and self-reported data across the perinatal period. The present analysis focused on 3 key features, including HRV, sleep (deep sleep and awake time), and fatigue, drawn from multimodal data collected through devices like the Oura Ring (Oura Health Inc) [22]. These features were chosen to capture both objective and subjective measures of health and align with recently published maternal health wearable studies. Nighttime HRV was calculated using the root mean square of successive differences between normal heartbeats method from 5-minute interval data collected by the Oura Ring. Awake time and deep sleep were derived from the Oura Ring's 3D accelerometer measurements and 5-minute sleep rankings, respectively. Fatigue was a self-reported feature tracked daily via the BUMP study smartphone app.

Recruitment

Participants were recruited through systematic, multichannel digital outreach strategies as described in the published BUMP study protocol [21]. Recruitment occurred primarily via a secure patient-provider platform operated by Sema4, through which potentially eligible individuals received study invitations. Additional recruitment efforts included targeted outreach through digital advertisements and social media platforms to broaden reach and ensure diverse participation.

Ethical Considerations

The BUMP study was approved by the Advarra Institutional Review Board (Pro00047893). Electronic informed consent was obtained from all participants through the study application prior to any data collection. Participant privacy and confidentiality were protected through data deidentification and secure data storage practices. Access to coded study data is restricted and governed through controlled-access platforms in accordance with institutional and regulatory requirements. Participant compensation, where applicable, was provided as outlined in the approved study protocol.

Participants

Out of 431 participants, 275 were included after excluding those with more than 60% missing data, missing delivery information, or fewer than 30 data points. The number of participants in each complication group was as follows: 37 with preterm birth, 42 with gestational diabetes mellitus, 38 with preeclampsia, 66 with gestational hypertension, 39 with postpartum hemorrhage, 189 with major depression, and 17 were categorized as healthy individuals (as defined as normal weight [BMI<25], age<35 years, and no complications during pregnancy and postpartum). Further demographic information is provided in Table S4 in [Multimedia Appendix 1](#).

Data Preprocessing

Details on data collection and study measures are in [Multimedia Appendix 1](#). We used aggregated daily summaries from 5-minute interval sleep data, calculating total duration for each sleep stage (deep sleep, awake, and rapid eye movement) across the sleep period. Moreover, HRV was averaged over the entire sleep period. Duplicated time points for objective features were removed. Outliers were identified using the 0.05 and 0.95 quantiles estimated from a randomly selected subset of 100 participants from the study population; these quantile thresholds were then applied uniformly across all participants and time points, with values outside the range set to not a number. For quadratic analysis, data were smoothed by gestational week, and averages were computed. Data preprocessing involved calculating the number of days relative to each participant's delivery date (days since delivery) at every sample point and determining the corresponding gestational age. Gestational age was calculated based on the estimated due date provided during the initial enrollment and adjusted according to the actual birth date when necessary. This information was crucial in aligning wearable device data and survey responses with the appropriate stage of pregnancy for accurate longitudinal analysis. Fatigue was kept in its continuous form between 1 and 7 values where 1 is "none" and 7 is "severe."

A case study analysis examined factors contributing to heterogeneity and varying signal behavior among participants, including self-reported severe daily events, severe symptom events, and adverse event (AE) reports. Severe symptoms, outlined in Table S2 in [Multimedia Appendix 1](#), were reported biweekly. To enhance the evaluation of their impact on distribution change analysis, we applied the SentenceTransformer [23] model to generate embeddings

for daily and biweekly features, enabling a more precise estimation of severe symptom timestamps.

Expanding this analysis, we examined the impact of AEs on feature trajectories using a dataset subset with free-text descriptions recorded by engagement experts after biweekly check-ins. To ensure consistency and extract insights, we used GPT-4-turbo to parse and label the free-text data, identifying explicit dates (MM/DD/YYYY or YYYY/MM/DD) and inferring incident dates from relative time references (eg, "one week ago"). GPT-4-turbo was prompted to generate concise labels for AEs by recognizing key occurrences, such as medical procedures or symptoms, and condensing them into meaningful terms. For example, "She started to get a cold 2 days ago, sinus and ear pressure, nasal congestion, and cough" was labeled as "Infection," while "Participant went to the emergency room with body aches and a fever" was labeled "Fever and body aches." This approach enabled efficient and consistent categorization of AEs, ensuring high-quality labeled data for analysis. Further details, including the complete list of event labels and their corresponding report counts, are provided in Table S5 in [Multimedia Appendix 1](#). Additional information on the GPT model and its use is described in [Multimedia Appendix 1](#).

Statistical Analysis

To assess the model's performance, we calculated the coefficient of determination (R^2), which quantifies how well the fitted quadratic model explains the variance in the observed data. An R^2 value closer to 1 indicates a better fit and was estimated as follows:

$$R^2 = 1 - \frac{\sum (y_{\text{observed}} - y_{\text{fitted}})^2}{\sum (y_{\text{observed}} - \mu_{\text{observed}})^2}$$

where y_{observed} , μ_{observed} , and y_{fitted} represent the initial feature pattern, mean of observed data, and quadratic model output, respectively.

Kernel density estimation is a nonparametric method for estimating the probability density function of a continuous variable, providing a smooth data distribution representation. Kernel density estimation places a Gaussian kernel on each data point and sums them up to generate a continuous density estimate, avoiding parametric assumptions. To compare probability density function distributions between subgroups, we applied the Kolmogorov-Smirnov test [24]. A P value <.05 indicates a statistically significant difference, rejecting the null hypothesis.

To evaluate the impact of including age as a predictor, we performed a likelihood ratio (LR) test comparing the full model with age to a reduced model without it. The log-likelihoods of both models were computed, denoted as ll_{full} for the full model and ll_{reduced} for the reduced model. The LR statistic was calculated as follows:

$$LR = -2 \times (ll_{\text{reduced}} - ll_{\text{full}})$$

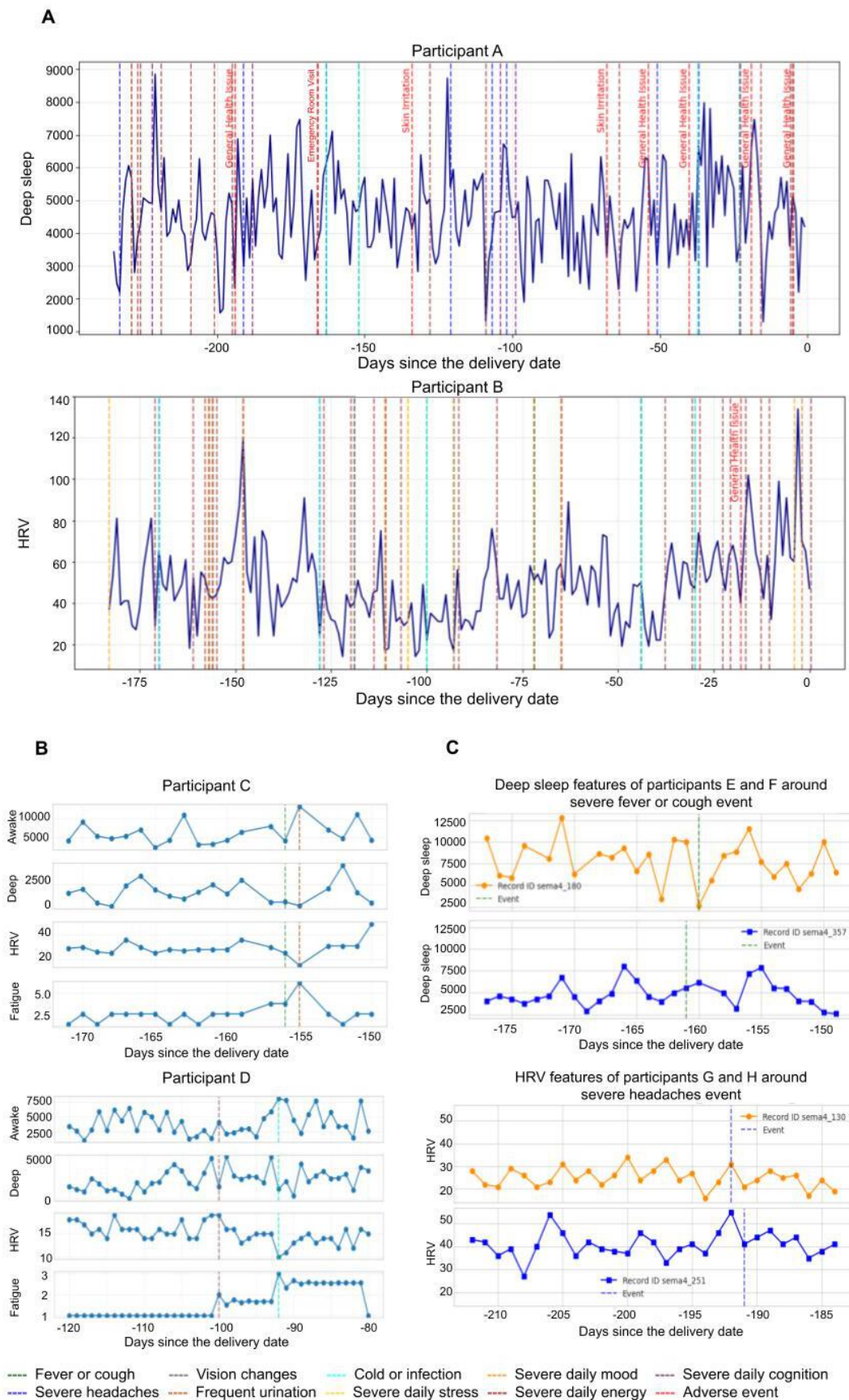
The degrees of freedom (df) for the χ^2 distribution were determined by the difference in the number of parameters between the full and reduced models. Using the χ^2 distribution, we computed the P value [25] based on the LR statistic and df , providing statistical evidence for comparing the 2 models.

For the case study analysis, we first conducted an intraindividual assessment. For each event, we compared signal behavior during a peri-event window (7 d before and after the event, defined relative to gestational age) to a baseline period (1 mo-7 d prior). Comparisons were included only if both the baseline and peri-event periods contained at least 7 data points, ensuring sufficient data for statistical analysis. We applied an unpaired Welch t test [26], which does not assume equal variance, to evaluate the null hypothesis of equal means between the 2 periods. The resulting t statistic and P value quantified the magnitude and statistical

significance of differences in signal behavior associated with the event window relative to baseline.

We then performed an interindividual analysis to assess whether similar patterns were observed across individuals. Using consecutive 14-day windows spanning gestational ages from -270 days to delivery (0), we identified the occurrence of each event within each window and calculated the average t statistic for each individual. When multiple events occurred in close temporal proximity, events were analyzed only if their peri-event windows did not overlap; otherwise, they were treated as a single composite event. For Figure 1B, the sampling frame consisted of all participants with at least 1 adverse or severe event for which the intraindividual Welch t test indicated a statistically significant difference ($P < .05$) between baseline and peri-event windows. From this subset, 1 qualifying event per participant was randomly selected using uniform random sampling without replacement.

Figure 1. Impact of adverse events and severe symptoms on fatigue, HRV, and sleep features over pregnancy. (A) Effects of adverse or severe events on deep sleep and HRV in example BUMP participants, with adverse events shown in red next to dashed lines. (B) Examples of adverse or severe events associated with significant shifts in the distribution of features throughout pregnancy. (C) Two women experiencing the same severe symptom at the same gestational week exhibiting different sleep and HRV patterns. BUMP: Better Understanding the Metamorphosis of Pregnancy; HRV: heart rate variability.



To quantify heterogeneity in responses, we defined a high-variability threshold as twice the SD of the average t statistic across individuals. If the absolute difference in t statistics between any two individuals exceeded this threshold, the event was classified as demonstrating substantial interindividual variability in associated feature changes. For [Figure 1C](#), the sampling frame included participants who experienced the same severe symptom within the same 14-day gestational window and demonstrated statistically significant physiological changes associated with that symptom. From this eligible pool, 2 participants were randomly selected using uniform random sampling to illustrate examples of interindividual variability in physiological response patterns.

Results

Overview of Analysis Strategy

We performed the analysis in 4 stages. First, we replicated previously reported aggregate-level maternal wearable findings in the independent BUMP cohort, focusing on HRV, fatigue, and sleep features. Second, we quantified the extent to which these aggregate patterns reflected individual-level behavior by examining participant-specific inflection points, extrema, slopes, and longitudinal trajectories. Third, we tested whether interpersonal heterogeneity could be explained by demographic or pregnancy-complication subgroups using distributions of individual model fit. Fourth, we conducted event-centered case studies to

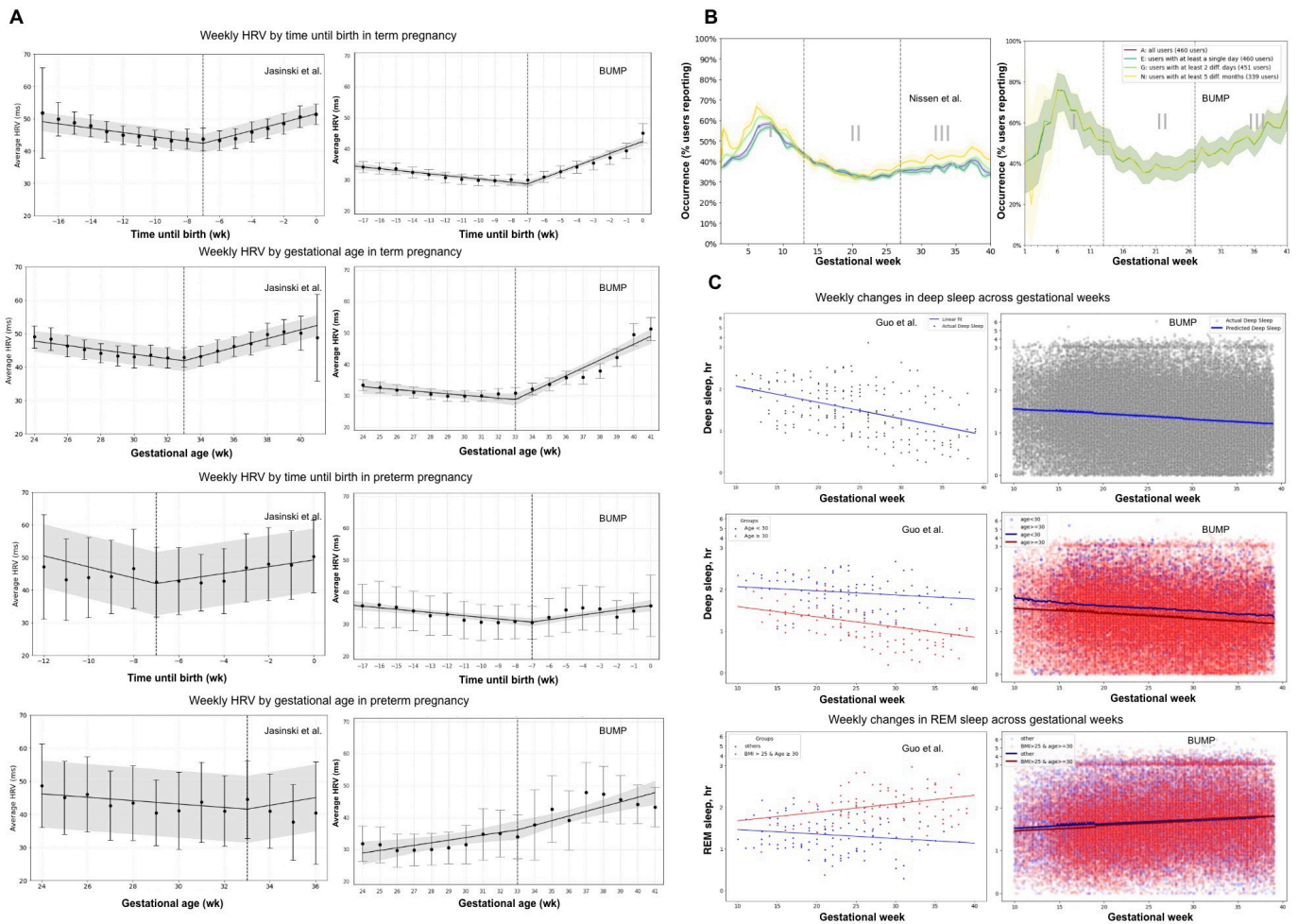
assess whether severe symptoms and AEs could account for deviations in individual trajectories.

Aggregated Analysis

To assess the generalizability of previously reported aggregate wearable-derived patterns in maternal health [16-18], we replicated analyses of HRV, fatigue, and sleep physiology using a larger, independent wearable dataset from the BUMP study ($n=256$) [21]. Aggregate trajectories over gestation were compared with those reported in prior studies, including the timing of inflection points and overall directional trends. In the following, we describe the degree to which these previously observed patterns were reproduced in the BUMP cohort and highlight areas of concordance and divergence across outcomes.

In the BUMP study, the aggregated results suggest that nighttime daily HRV shows a similar inflection point around week 7 before birth in both term and preterm pregnancies and at gestational week 33 in term pregnancies, though this trend is absent in preterm pregnancies ([Figure 2A](#)). However, the weekly mean HRV values differ significantly, with the BUMP data showing a smaller weekly variance compared to Jasinski et al [16]. For fatigue, the general trends in the BUMP study reproduce those reported in Nissen et al [17], with both showing an early peak around gestational weeks 7 to 8 and a trough around week 21, although substantial deviations from this pattern are observed at the individual level, particularly in the third trimester ([Figure 2B](#)).

Figure 2. Aggregated results comparison of HRV, fatigue, and sleep patterns between BUMP [21] and prior pregnancy studies. (A) HRV trends show similar prebirth inflections, with less variance in the BUMP study. Error bars represent the 95% CI; the vertical dashed line marks the inflection point [16]. (B) Fatigue aggregate peaks and the troughs pattern in the BUMP study reproduce previously reported population-level trends, despite substantial variability in individual trajectories. Dashed vertical lines mark the pregnancy trimesters [17]. (C) Deep sleep trends differ in slope with less variation across BMI and age subgroups in the BUMP study [18]. BUMP: Better Understanding the Metamorphosis of Pregnancy; HRV: heart rate variability; REM: rapid eye movement.

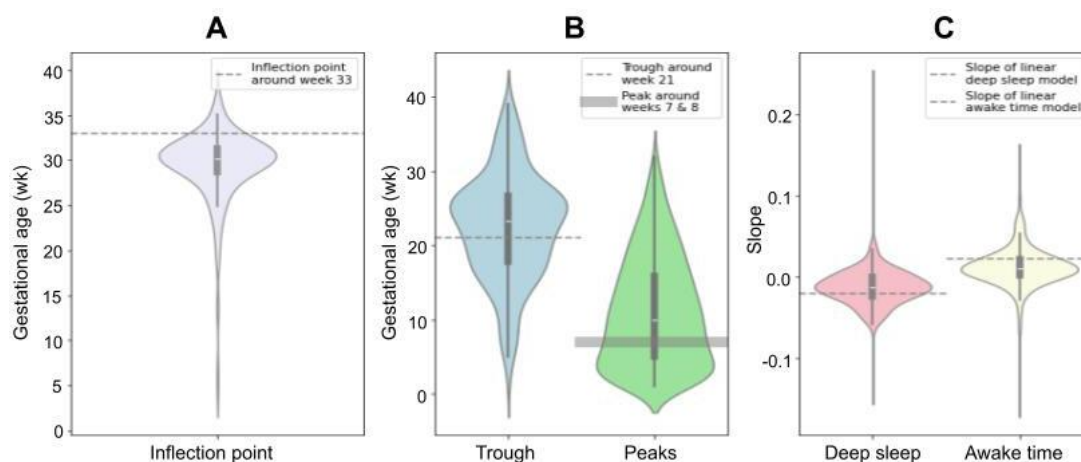


Regarding deep sleep (Figure 2C), both the BUMP data and Guo et al.'s findings show a negative slope for aggregate deep sleep, with Guo et al. [18] reporting a slightly steeper decline across 10 to 40 gestational weeks. Our analysis further reveals substantial variability and heterogeneity in sleep data, reflected in intraclass correlation coefficients of 0.618 for deep sleep and 0.502 for rapid eye movement sleep, both exceeding the 0.5 threshold for significance. In the BUMP data, BMI and age were not clearly separable predictors, unlike Guo et al.'s findings. Segmenting age into ≥ 30 years versus < 30 years did not significantly improve model fit ($P=.17$), suggesting that Guo et al.'s proposed associations may not be generalizable to all populations. Similarly, stratifying the data by prepregnancy BMI and age (age ≥ 30 and BMI > 25 vs others) yielded a P value of 1, further questioning the generalizability of these demographic

factors as predictive indicators. These findings highlight that results from wearable maternal health studies are not easily generalizable across populations due to significant heterogeneity in wearable feature patterns and individual subjective symptom experiences.

The analysis of HRV, fatigue, and sleep metrics derived from personal DHT data during pregnancy underscores the limitations of aggregated models in capturing individual variability in maternal health. Figure 3 shows violin plots, using density curves to illustrate data distribution. Wider sections indicate higher data density, enabling intuitive comparisons of individual variability against aggregate trends. Gray lines mark key findings from each study, contrasted with BUMP dataset analyses to assess density across individual-level data.

Figure 3. Key features extracted from wearable data during pregnancy in the BUMP study: (A) HRV, (B) fatigue, and (C) sleep. Horizontal dashed or highlighted lines represent the findings from Jasinski et al [16], Nissen et al [17], and Guo et al [18] based on aggregate results. The violin plots compare these aggregate findings with individual-level BUMP analysis, highlighting variability across participants. BUMP: Better Understanding the Metamorphosis of Pregnancy; HRV: heart rate variability.



Key comparisons include the following: For HRV, Jasinski et al [16] identified an inflection point around gestational week 33. To compare with individual-level inflection points, the BUMP data were analyzed using a quadratic model (weeks 10-40), identifying the minimum point as the individual inflection. Similarly, aggregate fatigue trends in BUMP are consistent with Nissen et al [17], showing early peaks at weeks 7 to 8 and a trough near week 21. Individual fatigue peaks and troughs were calculated to capture variability. For sleep, Guo et al [18] used a linear mixture model to estimate slopes for aggregated deep sleep and awake time, which were compared with individual-level slopes determined by fitting linear regression models.

Figure 3 and Table S7 in Multimedia Appendix 1 show that individual-level results from BUMP, visualized through

density curves, did not fully align with the published findings and revealed substantial variation. Figure 3B highlights that the median value of the gestational week in individual fatigue troughs was 23 (IQR 8) weeks, which was 2 weeks earlier than the aggregate result. Similarly, individual fatigue peaks occurred 2 weeks later than the aggregate finding, with notable variability indicated by the IQR and density spread. These discrepancies underscore the need to account for individual variability, as aggregate findings do not universally apply, even to the majority of individuals.

N-of-1 Level Analysis

Individual patterns in HRV, fatigue, deep sleep, and awake time were examined and compared to aggregate trends using the BUMP study data (Figure 4).

Figure 4. Spaghetti plots of four features from the BUMP study contrast aggregate trends with individual trajectories. Each colored line represents an individual's (A) HRV, (B) fatigue, (C) deep sleep time, and (D) awake time, highlighting variability across gestational weeks. BUMP: Better Understanding the Metamorphosis of Pregnancy; HRV: heart rate variability.

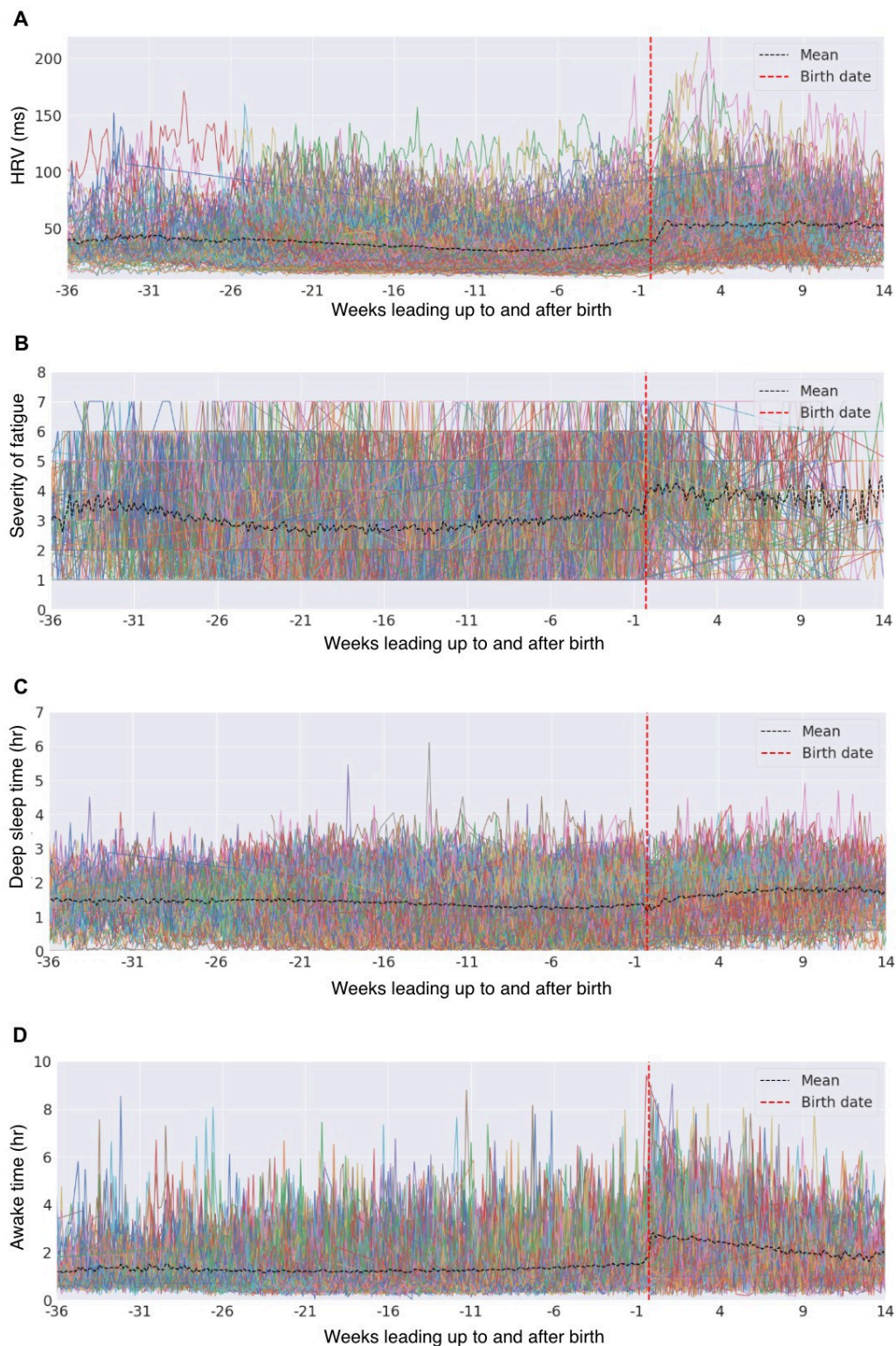


Figure 4 highlights the high variability in objective and subjective feature patterns during pregnancy, often lost in aggregate views. Individual HRV, fatigue, deep sleep, and awake time patterns vary significantly across gestational weeks. While the aggregate line (black) suggests a smooth trend, it overlooks individual complexities and key variations.

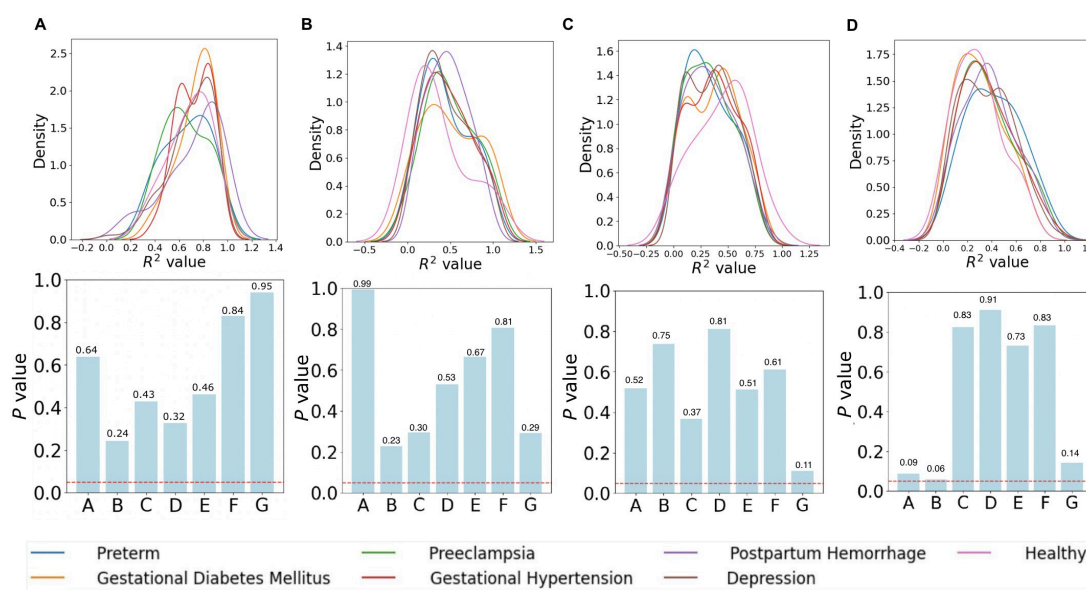
Relying solely on aggregate data risks misinterpretation, especially when individual patterns deviate from the average.

In Figure 5, heterogeneity is formally quantified using distributions of individual-level model fit (R^2) and tested across subgroups using kernel density comparisons with associated P values. Modifying factors or confounders, such as pregnancy-related complications and demographic factors,

may in part explain the observed heterogeneity in the model's performance. However, the kernel density plots and P values revealed no significant differences in R^2 distributions, indicating that the model's fit is relatively consistent across groups. Specifically, there were no significant differences in HRV, fatigue, deep sleep, or awake time trends among participants with different pregnancy-related conditions. All subgroup comparisons yielded P values $>.05$ (eg, $P=.09$, $.06$, $.82$ for awake time; $P=.52$, $.75$, $.37$ for deep sleep; $P=.99$, $.23$,

$.30$ for fatigue; and $P=.64$, $.24$, $.43$ for heart rate variability) (see the lower section of Figure 5). The symmetric, overlapping kernel density plots suggest minimal variation in R^2 distributions, with no skewness or distinct clustering. These findings suggest that pregnancy-related complications and demographic factors (Figure S1 in Multimedia Appendix 1) were not associated with detectable differences in the model's ability to fit the observed outcomes.

Figure 5. Comparison of R^2 values assessing model fit across pregnancy-related complications in the BUMP study: (A) HRV, (B) fatigue, (C) deep sleep, and (D) awake time. Top plots illustrate kernel density distributions for each group, while bottom plots display P values, comparing each group's feature distribution to that of other women. The red-dashed line marks the significance threshold ($P=.05$). Kernel density plots and P values indicate no statistically significant differences detected in R^2 distributions across complication subgroups. BUMP: Better Understanding the Metamorphosis of Pregnancy; HRV: heart rate variability.



Case Studies

To explore sources of heterogeneity, we analyzed survey data on severe symptoms reported by users and AEs documented in research staff notes. Key events and their precise timing were identified from unstructured AE notes using GPT-based automatic prompting, with subsequent expert review to verify the accuracy and consistency of the extracted information. We also incorporated daily severe symptoms: very low energy, impaired cognition, very low mood, and extreme stress, referred to as severe daily features. Biweekly severe symptom surveys, aligned with daily reports, covered severe fever or cough, intense headaches, severe infections, vision changes, and urinary issues. Further details are available in the *Methods* section and Multimedia Appendix 1.

Figure 1A presents examples of high-frequency AEs and severe symptoms reported during pregnancy. Among 256 individuals, 4674 events were recorded, with a median of 9 (IQR 8) and an average of 21 (SD 25.337; range 1-153) events per person. Figure 1B highlights 2 participants randomly selected from those with statistically significant shifts in feature distributions before and after severe events, as determined by an intraindividual Welch t test comparing baseline and peri-event periods ($P<.05$). In participant C, a low-energy event 155 days before delivery was followed by

increased fatigue and awake time, along with decreased HRV and deep sleep. In participant D, a severe infection 92 days before delivery was followed by similar effects.

Figure 1C illustrates variability in physiological patterns among pregnant women experiencing the same severe symptom at the same gestational week. Participants were randomly selected from those with significant physiological responses to the same symptoms, highlighting individual variability in adaptation. For example, both participants E and F experienced severe fever, sore throat, and/or cough in the same week; yet one had a sharp decline in deep sleep, whereas the other showed a gradual increase. A linear mixed model analysis indicated substantial between-individual variability in responses across all event types. These findings underscore the need to consider both inter- and intraindividual differences when evaluating responses to similar events. Additional examples are shown in Figure S4 in Multimedia Appendix 1.

Discussion

Principal Findings

This study aimed to evaluate whether aggregate trends in DHT measures of HRV, sleep, and self-reported fatigue

during pregnancy accurately represent individual experiences and to explore whether demographic or clinical factors explain variability among individuals. Using the BUMP dataset and N-of-1 spline modeling, we report 3 principal findings. First, aggregate trends in HRV, sleep, and fatigue were broadly consistent with those reported in prior studies but masked substantial interindividual variability. Second, N-of-1 analyses revealed that many participants deviated markedly from population averages in the timing, direction, and magnitude of changes across pregnancy. Third, neither demographic factors, pregnancy complications, nor acute health events adequately explained the observed heterogeneity, suggesting that personalized analytical approaches are necessary to accurately interpret maternal health data.

Comparison to Prior Work

The aggregate results from these studies and the BUMP dataset demonstrate unique contours in these 3 measures of health. Notably, HRV appears to steadily decrease with gestational age, but then shows an inflection at approximately 33 weeks leading up to delivery. This inflection has been highlighted in the literature by others, suggesting a phenomenon that could reflect an individual's physiological readiness for delivery [16,27,28]. While this could be true for some, our N-of-1 analyses demonstrate that many individuals do not experience this HRV inflection, or an inflection is experienced, but in a significantly different pattern from the average. Fatigue aggregate trends show an early peak around gestational weeks 7 to 8, followed by a gradual decline to a trough near week 21 [17]. However, individual analyses reveal considerable variations in these patterns, as shown in Figure 3. Similarly, while average sleep trends show a steady decline in deep sleep and increased awake time, especially in the third trimester [18,29], N-of-1 analyses highlight substantial interindividual differences. Some individuals experience a progressive decline in deep sleep, whereas others show minimal or even opposite changes, underscoring the complexity of sleep adaptations during pregnancy.

Physiological and subjective changes during pregnancy often differ from group averages due to external factors. While previous studies [5,18] have often relied on standard demographic exclusions or subgrouping and stratified analyses by age or pregnancy complications, these approaches miss the substantial heterogeneity in individual patterns. Using N-of-1 analyses with spline models, we captured each participant's unique response to pregnancy-related changes, offering detailed insights into individual trajectories. Our N-of-1 analyses found no significant differences in HRV, sleep, and fatigue levels between participants with pregnancy complications and those without, suggesting that these conditions do not fully explain the observed heterogeneity. Some differences emerged across demographic factors, such as age and BMI; for instance, individuals with a BMI of 25 to 30 showed significant differences in HRV and deep sleep goodness-of-fit compared to other subgroups. However, this pattern was not observed for fatigue or awake time, indicating BMI alone is not a reliable indicator for subgrouping but warrants further exploration alongside other factors.

We further explored case studies in an attempt to isolate potential modifying factors of the examined maternal health metrics. The case studies further compounded the finding suggesting high heterogeneity of maternal health experiences. For example, in some participants, while there were associations between events, such as fever, cough, and very low energy levels, and significant increases in fatigue and awake time, there was high intraindividual and interindividual variability in responses to these events, emphasizing that pregnant individuals can exhibit diverse physiological responses to the same event at the same gestational week.

After extensive exploration into potential modifiers of HRV, sleep, and fatigue over pregnancy, we identify no clear factors explaining the observed high heterogeneity of these maternal health metrics, suggesting the necessity of personalized analytical approaches to accurately interpret maternal health data at the N-of-1 level. The significant variability among individuals highlights the complexity of deriving broadly applicable clinical insights. The absence of significant associations with the examined demographic or clinical factors should be interpreted cautiously, as limited power in smaller subgroups and unmeasured factors may contribute to the observed variability.

Limitations

Several limitations should be considered when interpreting these findings. First, the sample size limited statistical power for subgroup analyses; small subgroup sizes may have prevented detection of meaningful associations between demographic or clinical factors and the observed variability. Second, multiple unmeasured confounders may contribute to the heterogeneity observed and were not captured in this study. Third, we note that those prior studies and BUMP differed slightly in cohort selection criteria and protocol, but such differences are unlikely to affect our conclusion, as the studies used the same device types, measurement protocols, and aggregate feature definitions for comparison.

Conclusions

This study underscores the need for personalized and diverse assessments to capture the individuality of maternal health metrics. Moving beyond one-size-fits-all models is essential for embracing heterogeneity and tailoring insights to pregnant individuals' unique needs. By explicitly examining N-of-1 longitudinal patterns, this study contributes to early evidence highlighting the methodological considerations necessary for safe, interpretable, and personalized maternal health monitoring. Future work should prioritize developing personalized methods that account for variability and assess the causal effects of modifying factors. Effectively interpreting personal DHT maternal health data may require human-in-the-loop approaches, where users are prompted for additional context when deviations in feature patterns are detected.

Looking forward, the findings of this study motivate the development of analytical frameworks that explicitly accommodate individual variability while leveraging population-level information. A clinically actionable N-of-1

framework would require models that explicitly borrow statistical strength from population-level data while producing calibrated, individualized inferences. One promising direction is hierarchical Bayesian time-series modeling, in which population-level dynamics inform individual trajectories without enforcing a shared average response. For example, hierarchical Dirichlet process flow-style approaches [30,31] allow each person to follow a personalized latent

dynamic while sharing structure across the cohort, improving individual-level prediction under sparse or noisy observations. In pregnancy monitoring, such models could learn typical gestational regimes while estimating patient-specific deviations, enabling individualized forecasting (eg, expected HRV trajectory) and anomaly detection relative to one's own baseline rather than a cohort mean.

Acknowledgments

The authors thank Phoebe Chuang for her assistance with data analysis. They would also like to thank all the women who participated in the BUMP (Better Understanding the Metamorphosis of Pregnancy) and BUMP-C studies for their ongoing commitment and valuable contributions to enhancing future pregnancy care. The authors declare the use of generative AI (GAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GAI tools under full human supervision: proofreading and editing. The GAI tool used was ChatGPT-4o (OpenAI). Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed by the authors and do not bear responsibility for the final outcomes.

Funding

This work was partially supported by Sema4 and 4YouandMe, a nonprofit organization based in Seattle, Washington, United States, with additional in-kind contributions from coalition partners including Evidation Health, Vector Institute, Cambridge Cognition, and Bodyport. Additional support was provided by the Varma Family Chair (AG) and CIFAR AI Chair (AG). TB was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) doctoral award. TB and JY were supported by the Vector Scholarship from the Vector Institute and Ontario Graduate Scholarship from the Province of Ontario. The funding sources played no role in study design or analysis.

Data Availability

Among participants who opt-in, the coded study data from the BUMP study participants are available on the Synapse platform at Sage Bionetworks [32] and can be freely accessed by any researcher who becomes "qualified" by becoming a registered and certified Synapse user [33], and by meeting the specific conditions of use that require submitting an intended data use statement alongside an Institutional Review Board (IRB)-approved protocol. The BUMP specific Synapse Project page [34] is listed as references and is accessible to registered Synapse users. The details of Reproducibility GitHub link [35] and the website page [36] are listed as references.

Authors' Contributions

JY and TB designed the methodology, conducted formal analysis, developed the software, interpreted and validated the results, and drafted the original manuscript. RY contributed to methodology design, analysis, software development, result interpretation, validation, data curation, and manuscript review and editing. AB assisted with data curation, software development, analysis, and manuscript review and editing. AG and SMG contributed to conceptualization, supervised the analysis, interpreted results, and reviewed and edited the manuscript. SHF led the overall study design and conceptualization, secured funding, supervised the analysis, interpreted results, and reviewed and edited the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Detailed methodology, data collection procedures, and extended analyses.

[DOCX File (Microsoft Word File), 6300 KB-Multimedia Appendix 1]

Checklist 1

STROBE checklist.

[PDF File (Adobe File), 129 KB-Checklist 1]

References

1. Vogel JP, Jung J, Lavin T, et al. Neglected medium-term and long-term consequences of labour and childbirth: a systematic analysis of the burden, recommended practices, and a way forward. *Lancet Glob Health*. Feb 2024;12(2):e317-e330. [doi: [10.1016/S2214-109X\(23\)00454-0](https://doi.org/10.1016/S2214-109X(23)00454-0)] [Medline: [38070535](https://pubmed.ncbi.nlm.nih.gov/38070535/)]
2. Jiang L, Tang K, Magee LA, et al. A global view of hypertensive disorders and diabetes mellitus during pregnancy. *Nat Rev Endocrinol*. Dec 2022;18(12):760-775. [doi: [10.1038/s41574-022-00734-y](https://doi.org/10.1038/s41574-022-00734-y)] [Medline: [36109676](https://pubmed.ncbi.nlm.nih.gov/36109676/)]

3. Sharifi-Heris Z, Rahmani AM, Axelin A, Rasouli M, Bender M. Heart rate variability and pregnancy complications: systematic review. *Interact J Med Res*. Jun 5, 2023;12:e44430. [doi: [10.2196/44430](https://doi.org/10.2196/44430)] [Medline: [37276013](https://pubmed.ncbi.nlm.nih.gov/37276013/)]
4. Goodday SM, Karlin E, Alfarano A, et al. Correction: an alternative to the light touch digital health remote study: the stress and recovery in frontline COVID-19 health care workers study. *JMIR Form Res*. Apr 18, 2022;6(4):e38188. [doi: [10.2196/38188](https://doi.org/10.2196/38188)] [Medline: [35436210](https://pubmed.ncbi.nlm.nih.gov/35436210/)]
5. Modde Epstein C, McCoy TP. Linking electronic health records with wearable technology from the All of Us research program. *J Obstet Gynecol Neonatal Nurs*. Mar 2023;52(2):139-149. [doi: [10.1016/j.jogn.2022.12.003](https://doi.org/10.1016/j.jogn.2022.12.003)] [Medline: [36702164](https://pubmed.ncbi.nlm.nih.gov/36702164/)]
6. Keeler Bruce L, González D, Dasgupta S, Smarr BL. Biometrics of complete human pregnancy recorded by wearable devices. *NPJ Digit Med*. Aug 12, 2024;7(1):207. [doi: [10.1038/s41746-024-01183-9](https://doi.org/10.1038/s41746-024-01183-9)] [Medline: [39134787](https://pubmed.ncbi.nlm.nih.gov/39134787/)]
7. Goodday SM, Geddes JR, Friend SH. Disrupting the power balance between doctors and patients in the digital era. *Lancet Digit Health*. Mar 2021;3(3):e142-e143. [doi: [10.1016/S2589-7500\(21\)00004-2](https://doi.org/10.1016/S2589-7500(21)00004-2)] [Medline: [33516653](https://pubmed.ncbi.nlm.nih.gov/33516653/)]
8. Bhaltadak V, Ghewade B, Yelne S. A comprehensive review on advancements in wearable technologies: revolutionizing cardiovascular medicine. *Cureus*. May 2024;16(5):e61312. [doi: [10.7759/cureus.61312](https://doi.org/10.7759/cureus.61312)] [Medline: [38947726](https://pubmed.ncbi.nlm.nih.gov/38947726/)]
9. Liao Y, Thompson C, Peterson S, Mandrola J, Beg MS. The future of wearable technologies and remote monitoring in health care. *Am Soc Clin Oncol Educ Book*. Jan 2019;39:115-121. [doi: [10.1200/EDBK_238919](https://doi.org/10.1200/EDBK_238919)] [Medline: [31099626](https://pubmed.ncbi.nlm.nih.gov/31099626/)]
10. Erickson EN, Gotlieb N, Pereira LM, Myatt L, Mosquera-Lopez C, Jacobs PG. Predicting labor onset relative to the estimated date of delivery using smart ring physiological data. *NPJ Digit Med*. Aug 19, 2023;6(1):153. [doi: [10.1038/s41746-023-00902-y](https://doi.org/10.1038/s41746-023-00902-y)] [Medline: [37598232](https://pubmed.ncbi.nlm.nih.gov/37598232/)]
11. Mageri A, Barchitta M, Agodi A. How wearable sensors can support the research on foetal and pregnancy outcomes: a scoping review. *J Pers Med*. Jan 26, 2023;13(2):218. [doi: [10.3390/jpm13020218](https://doi.org/10.3390/jpm13020218)] [Medline: [36836452](https://pubmed.ncbi.nlm.nih.gov/36836452/)]
12. Nagaraj S, Goodday S, Hartvigsen T, et al. Dissecting the heterogeneity of “in the wild” stress from multimodal sensor data. *NPJ Digit Med*. Dec 20, 2023;6(1):237. [doi: [10.1038/s41746-023-00975-9](https://doi.org/10.1038/s41746-023-00975-9)] [Medline: [38123810](https://pubmed.ncbi.nlm.nih.gov/38123810/)]
13. Ravindra NG, Espinosa C, Berson E, et al. Deep representation learning identifies associations between physical activity and sleep patterns during pregnancy and prematurity. *NPJ Digit Med*. Sep 28, 2023;6(1):171. [doi: [10.1038/s41746-023-00911-x](https://doi.org/10.1038/s41746-023-00911-x)] [Medline: [37770643](https://pubmed.ncbi.nlm.nih.gov/37770643/)]
14. Aktar SF, Upama PB, Ahamed SI. Leveraging technology to address women’s health challenges: a comprehensive survey. Presented at: 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC); Jul 2-4, 2024:941-949; Osaka, Japan. [doi: [10.1109/COMPSAC61105.2024.00130](https://doi.org/10.1109/COMPSAC61105.2024.00130)]
15. Guk K, Han G, Lim J, et al. Evolution of wearable devices with real-time disease monitoring for personalized healthcare. *Nanomaterials (Basel)*. May 29, 2019;9(6):813. [doi: [10.3390/nano9060813](https://doi.org/10.3390/nano9060813)] [Medline: [31146479](https://pubmed.ncbi.nlm.nih.gov/31146479/)]
16. Jasinski SR, Rowan S, Presby DM, Claydon EA, Capodilupo ER. Wearable-derived maternal heart rate variability as a novel digital biomarker of preterm birth. *PLoS One*. 2024;19(1):e0295899. [doi: [10.1371/journal.pone.0295899](https://doi.org/10.1371/journal.pone.0295899)] [Medline: [38295026](https://pubmed.ncbi.nlm.nih.gov/38295026/)]
17. Nissen M, Barrios Campo N, Flaucher M, et al. Prevalence and course of pregnancy symptoms using self-reported pregnancy app symptom tracker data. *NPJ Digit Med*. Oct 11, 2023;6(1):189. [doi: [10.1038/s41746-023-00935-3](https://doi.org/10.1038/s41746-023-00935-3)] [Medline: [37821584](https://pubmed.ncbi.nlm.nih.gov/37821584/)]
18. Guo Y, Xu Q, Dutt N, Kehoe P, Qu A. Longitudinal changes in objective sleep parameters during pregnancy. *Womens Health (Lond)*. 2023;19:17455057231190952. [doi: [10.1177/17455057231190952](https://doi.org/10.1177/17455057231190952)] [Medline: [37650368](https://pubmed.ncbi.nlm.nih.gov/37650368/)]
19. Pollet TV, Stulp G, Henzi SP, Barrett L. Taking the aggravation out of data aggregation: a conceptual guide to dealing with statistical issues related to the pooling of individual-level observational data. *Am J Primatol*. Jul 2015;77(7):727-740. [doi: [10.1002/ajp.22405](https://doi.org/10.1002/ajp.22405)] [Medline: [25810242](https://pubmed.ncbi.nlm.nih.gov/25810242/)]
20. Young-Lin N, Heneghan C, Liu Y, et al. Insights into maternal sleep: a large-scale longitudinal analysis of real-world wearable device data before, during, and after pregnancy. *EBioMedicine*. Apr 2025;114:105640. [doi: [10.1016/j.ebiom.2025.105640](https://doi.org/10.1016/j.ebiom.2025.105640)] [Medline: [40147048](https://pubmed.ncbi.nlm.nih.gov/40147048/)]
21. Goodday SM, Karlin E, Brooks A, et al. Better Understanding of the Metamorphosis of Pregnancy (BUMP): protocol for a digital feasibility study in women from preconception to postpartum. *NPJ Digit Med*. Mar 30, 2022;5(1):40. [doi: [10.1038/s41746-022-00579-9](https://doi.org/10.1038/s41746-022-00579-9)] [Medline: [35354895](https://pubmed.ncbi.nlm.nih.gov/35354895/)]
22. Altini M, Kinnunen H. The promise of sleep: a multi-sensor approach for accurate sleep stage detection using the Oura Ring. *Sensors (Basel)*. Jun 23, 2021;21(13):4302. [doi: [10.3390/s21134302](https://doi.org/10.3390/s21134302)] [Medline: [34201861](https://pubmed.ncbi.nlm.nih.gov/34201861/)]
23. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using siamese BERT-networks. Presented at: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); Nov 3-7, 2019; Hong Kong, China. [doi: [10.18653/v1/D19-1410](https://doi.org/10.18653/v1/D19-1410)]

24. Berger VW, Zhou Y. Kolmogorov-Smirnov Test: overview. In: StatsRef: Statistics Reference Online. John Wiley & Sons; 2014. [doi: [10.1002/9781118445112.stat06558](https://doi.org/10.1002/9781118445112.stat06558)] ISBN: 9781118445112
25. Altman J. Observational study of behavior: sampling methods. Behaviour. 1974;49(3):227-267. [doi: [10.1163/156853974x00534](https://doi.org/10.1163/156853974x00534)] [Medline: [4597405](https://pubmed.ncbi.nlm.nih.gov/4597405/)]
26. Welch BL. The generalisation of student's problems when several different population variances are involved. Biometrika. 1947;34(1-2):28-35. [doi: [10.1093/biomet/34.1-2.28](https://doi.org/10.1093/biomet/34.1-2.28)] [Medline: [20287819](https://pubmed.ncbi.nlm.nih.gov/20287819/)]
27. Sarhaddi F, Azimi I, Axelin A, Niela-Vilen H, Liljeberg P, Rahmani AM. Trends in heart rate and heart rate variability during pregnancy and the 3-month postpartum period: continuous monitoring in a free-living context. JMIR mHealth uHealth. Jun 3, 2022;10(6):e33458. [doi: [10.2196/33458](https://doi.org/10.2196/33458)] [Medline: [35657667](https://pubmed.ncbi.nlm.nih.gov/35657667/)]
28. Rowan SP, Lilly CL, Claydon EA, Wallace J, Merryman K. Monitoring one heart to help two: heart rate variability and resting heart rate using wearable technology in active women across the perinatal period. BMC Pregnancy Childbirth. Nov 30, 2022;22(1):887. [doi: [10.1186/s12884-022-05183-z](https://doi.org/10.1186/s12884-022-05183-z)] [Medline: [36451120](https://pubmed.ncbi.nlm.nih.gov/36451120/)]
29. Mindell JA, Cook RA, Nikolovski J. Sleep patterns and sleep disturbances across pregnancy. Sleep Med. Apr 2015;16(4):483-488. [doi: [10.1016/j.sleep.2014.12.006](https://doi.org/10.1016/j.sleep.2014.12.006)] [Medline: [25666847](https://pubmed.ncbi.nlm.nih.gov/25666847/)]
30. Tonekaboni S, et al. HDP-flow: generalizable bayesian nonparametric model for time series state discovery. Presented at: Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence (UAI); Jul 21-25, 2025:4227-4250; Rio de Janeiro, Brazil. URL: <https://openreview.net/forum?id=Ti1TMJjprQ> [Accessed 2025-10-05]
31. Schulam P, Saria S. A framework for individualizing predictions of disease trajectories by exploiting multi-resolution structure. Presented at: NIPS'15: Proceedings of the 29th International Conference on Neural Information Processing Systems; Dec 7-12, 2015; Montreal, Canada. URL: https://proceedings.neurips.cc/paper_files/paper/2015/file/285e19f20beded7d215102b49d5c09a0-Paper.pdf [Accessed 2026-03-31]
32. Sage Bionetworks. URL: <https://sagebionetworks.org> [Accessed 2026-04-13]
33. Synapse user account types. Synapse. URL: <https://docs.synapse.org/synapse-docs/synapse-user-account-types> [Accessed 2026-06-13]
34. Better understanding the metamorphosis of pregnancy (BUMP). Synapse. URL: <https://www.synapse.org/#!/Synapse:syn25953345/wiki/616547> [Accessed 2026-04-13]
35. Behrouzi T, Yu J, Yang R. N-of-1 level wearable data analysis. GitHub. URL: <https://github.com/Tinbeh97/N-of-1-Level-Wearable-Data-Analysis> [Accessed 2026-04-13]
36. Maternal health aggregated trends can be misleading. N-of-1 Level Wearable Data Analysis. URL: <https://tinbeh97.github.io/N-of-1-Level-Wearable-Data-Analysis/?token=n1wearable123> [Accessed 2026-04-13]

Abbreviations

AEs: adverse events

BUMP: Better Understanding the Metamorphosis of Pregnancy

DHT: digital health technology

HRV: heart rate variability

LR: likelihood ratio

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by Amaryllis Mavragani; peer-reviewed by Jinyu Guo, Lars Masannek; submitted 20.Oct.2025; final revised version received 17.Mar.2026; accepted 17.Mar.2026; published 28.Apr.2026

Please cite as:

Behrouzi T, Yu J, Yang R, Boch A, Goldenberg A, Goodday SM, Friend SH

Benefit of the N-of-1 Approach Versus Aggregate Analysis in Tracking Individual Trajectories During Pregnancy: Comparison of Longitudinal Wearable Observational Studies

JMIR Form Res 2026;10:e86203

URL: <https://formative.jmir.org/2026/1/e86203>

doi: [10.2196/86203](https://doi.org/10.2196/86203)

© Tina Behrouzi, Jennifer Yu, Robin Yang, Adrien Boch, Anna Goldenberg, Sarah M Goodday, Stephen H Friend. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 28.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.