INTACT: A method for integration of longitudinal physical activity data from multiple sources

Jingru Zhang*[1], Erjia Cui[2], Hongzhe Li[3], and Haochang Shou[3]

[1]School of Data Science, Fudan University, Shanghai, 200433, China
[2]Division of Biostatisitcs and Health Data Science, University of Minnesota, Minneapolis, MN 55414
[3]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA 19104

This manuscript has been accepted for publication in Biometrics.

Abstract

Wearable devices and digital phenotyping are increasingly used in observational and interventional studies to measure real-time biosignals such as physical activity. However, integrating and comparing data across studies and cohorts remains challenging due to variability in device types, acquisition protocols, and preprocessing methods. A key challenge is removing unwanted study- or device-specific effects while preserving meaningful biological signals. These difficulties are exacerbated by the longitudinal and within-day correlations inherent in high-resolution time-varying data collected from wearable sensors. To address this, we propose INTACT (INtegration of Time-varying data from weArable sensors for physiCal acTivity), a novel method for harmonizing time-varying physical activity intensity data from accelerometers. INTACT models shared information through common eigenvalues and eigenfunctions while allowing for source-specific scale and rotation adjustments. We apply the proposed method to two real-world applications: (i) integration of accelerometer data from two waves of the National Health and Nutrition Examination Survey (NHANES), measured using different devices and reported in different units; and (ii) integration of NHANES accelerometry data with accelerometer and gyroscope measures from commercial devices. Across both applications, INTACT outperforms existing approaches in mitigating source effects while preserving biological variation, enabling more reliable cross-study comparisons of physical activity patterns.

Key words: Harmonization; Longitudinal data; NHANES data; Physical activity; Source effects; Wearable device.

1 Introduction

The rapid proliferation of wearable sensor devices has enabled continuous monitoring of diverse physiological and behavioral signals over extended periods. However, as the technology landscape evolves—with frequent releases of new devices and continual updates to existing algorithms (Harrison et al., 2015)—substantial variability arises in how devices are selected and data are collected, making cross-study integration highly challenging. Such heterogeneity manifests in multiple ways. For example, an open-source device reporting raw acceleration in 'g' units may be worn on the hip in one study and on the

*jr_zhang@fudan.edu.cn

wrist in another; studies may use devices placed on the same wrist but from different models; or proprietary commercial-grade devices may report step counts using undisclosed and potentially inconsistent algorithms. Even when multiple sources measure the same or related biological signals and show broadly comparable patterns, source-specific characteristics can still differ markedly. A notable case is the National Health and Nutrition Examination Survey (NHANES): in 2003–2004, physical activity was recorded using the Actigraph AM-7164, which reported activity counts (mean = 222.2), whereas in 2013–2014, the Actigraph GT3X+ was used, producing the Monitor-Independent Movement Summary (MIMS) metric (mean = 8.863).

Despite the widespread use of wearable sensors, few methods address cross-source heterogeneity. By contrast, many fields have developed integration techniques for source or batch effects—systematic differences unrelated to biological signals. In this paper, we use "source" broadly to denote a site, device, or batch of homogeneous data. For example, Benito et al. (2004) applied distance-weighted discrimination to microarray data, while Johnson et al. (2007) proposed linear model adjustments for source-specific means and scales. Similar approaches have been adapted for multimodal neuroimaging, including longitudinal ComBat (Beer et al., 2020) and CovBat (Chen et al., 2022). Additionally, methods for integrating single-cell data across multiple sites have been developed, including canonical correlation analysis (Butler et al., 2018), mutual nearest neighbors (Hie et al., 2019), and non-negative matrix factorization (Welch et al., 2019; Peng et al., 2021).

However, the longitudinal and high-frequency nature of physical activity data poses challenges for existing integration methods. Most either accommodate only a single measurement per participant or, when applied to longitudinal data, adjust mean and variance for each dimension separately, overlooking joint temporal structure and cross-dimensional correlations essential for integrating physical activity data. To address the longitudinal and within-day correlations in physical activity data, we propose INTACT (INtegration of Time-varying data from weArable sensors for physiCal acTivity), a framework for harmonizing accelerometer data across studies while accounting for source-specific variation. INTACT addresses the challenge of repeated measures over multiple days by separately modeling between- and within-subject variations and applying source effect corrections accordingly. It assumes a common eigenspace across datasets while allowing for source-specific shifts, rotations, and scale differences. This shared eigenspace captures consistent eigen-directions and magnitudes across datasets, effectively distinguishing biological signals from unwanted variability introduced by differences in data acquisition and preprocessing. By leveraging these shared components, INTACT mitigates source effects while preserving essential biological signals, enabling more reliable cross-study comparisons of physical activity patterns.

INTACT respects the inherent structure of epoch-level time-varying data from wearable devices, explicitly modeling repeated observations while preserving the original data distribution. Through an interpretable transformation, INTACT mitigates source effects without compromising the integrity of biological signals. For efficient implementation, we develop an iterative eigendecomposition algorithm for parameter estimation and provide the R package `intactPA` to facilitate application of the proposed method.

As motivated by the NHANES data integration example, our method is tailored to accelerometer data integration. Nevertheless, INTACT could potentially be applied to other longitudinal time-varying data requiring harmonization, such as continuous glucose monitoring or gyroscope data, provided that preprocessing and functional regression model specification are guided by prior knowledge (Sergazinov et al., 2023).

We demonstrate INTACT in two applications. The first integrates NHANES accelerometer data, collected using different devices across two study waves: Actigraph AM-7164 (NHANES 2003–2004) and Actigraph GT3X+ (NHANES 2013–2014). After harmonization, the two datasets exhibit improved consistency in activity intensity readings, a more pronounced age effect on activity patterns, and a stronger

association between physical activity and mortality. The second integrates NHANES 2013–2014 accelerometer data with accelerometer and gyroscope measures from commercial devices, yielding more consistent harmonized data and revealing a stronger association between sleep onset and next-day physical activity.

The remainder of this paper is organized as follows. Section 2 presents INTACT, our method for integrating longitudinal time-varying physical activity data. Section 3 evaluates its performance through simulations. Sections 4 and 5 illustrate its application in two real-world datasets. Section 6 concludes with a discussion.

2 Methodology for data integration

To integrate longitudinal physical activity data from heterogeneous sources, we propose INTACT, a method designed to adjust for source-specific variations while preserving biological signals. Section 2.1 introduces the model, Section 2.2 outlines the estimation and implementation steps, and Section 2.3 discusses strategies for incomplete data.

2.1 Model for integrating longitudinal physical activity data

We define a data source as a collection of observations acquired under consistent conditions, such as within a study or using the same device type. Suppose data are collected at the epoch level from M sources. Source a includes n_a subjects, each with r_i observed days, yielding $N_a = \sum_{i=1}^{n_a} r_i$ total observations. Let $Y_{ij}^{(a)}(t)$ denote the log-transformed activity intensity for subject i, day j, at time t from source a. To standardize scale across sources, we divide each $Y_{ij}^{(a)}(t)$ by the square root of the spectral norm of the global sample covariance matrix estimated for each source separately, as detailed in Step 1 of Section 2.2.

We model the scaled activity data $\tilde{Y}_{ij}^{(a)}(t)$ using the functional regression framework

$$\tilde{Y}_{ij}^{(a)}(t) = \alpha(t) + \gamma^{(a)}(t) + (Z_{ij}^{(a)})^T \beta(t) + \eta_{ij}^{(a)}(t), \quad a = 1, \ldots, M, \tag{1}$$

for subjects $i = 1, \ldots, n_a$ and replicates $j = 1, \ldots, r_i$, where $\alpha(t)$ denotes the overall mean function, $\gamma^{(a)}(t)$ is a source-specific mean shift, $Z_{ij}^{(a)}$ is a vector of biological covariates (e.g., age, sex), and $\beta(t)$ represents their time-varying effects. The residual term $\eta_{ij}^{(a)}(t)$ captures individual-level variation and is further decomposed into the structured functional model as considered in Shou et al. (2015) and Cui et al. (2023)

$$\eta_{ij}^{(a)}(t) = X_i^{(a)}(t) + W_{ij}^{(a)}(t) + \epsilon_{ij}^{(a)}(t), \tag{2}$$

where $X_i^{(a)}(t)$ and $W_{ij}^{(a)}(t)$ are uncorrelated, mean-zero subject-level and day-level latent processes with covariances $S^{X(a)}(s,t)$ and $S^{W(a)}(s,t)$, respectively. The term $\epsilon_{ij}^{(a)}(t)$ represents white noise measurement error. For identifiability, we impose $\sum_{a=1}^{M} N_a \gamma^{(a)}(t) = 0$, following Beer et al. (2020). We approximate the latent processes using truncated Karhunen–Loève expansions:

$$X_i^{(a)}(t) \approx \sum_{k=1}^{K_X} \sqrt{\lambda_k^{X(a)}} \psi_k^{X(a)}(t) \xi_{ik}^{X(a)}, \quad W_{ij}^{(a)}(t) \approx \sum_{k=1}^{K_W} \sqrt{\lambda_k^{W(a)}} \psi_k^{W(a)}(t) \xi_{ijk}^{W(a)}, \tag{3}$$

where $\psi_k^{X(a)}(t)$ and $\psi_k^{W(a)}(t)$ are eigenfunctions of $S^{X(a)}(s,t)$ and $S^{W(a)}(s,t)$, respectively, $\lambda_k^{X(a)}$, $\lambda_k^{W(a)}$ are the corresponding eigenvalues, and $\xi_{ik}^{X(a)}$, $\xi_{ijk}^{W(a)}$ are independent standard normal scores. The integers K_X and K_W specify the number of retained components.

Let $\Psi^{X(a)}(t) = (\psi_1^{X(a)}(t), \ldots, \psi_{K_X}^{X(a)}(t))$, $\Delta^{X(a)} = \text{diag}(\lambda_1^{X(a)}, \ldots, \lambda_{K_X}^{X(a)})$, and $\xi_i^{X(a)} = (\xi_{i1}^{X(a)}, \ldots, \xi_{iK_X}^{X(a)})^T$, with analogous definitions for the W part. Then, (3) can be expressed as

$$X_i^{(a)}(t) \approx \Psi^{X(a)}(t)(\Delta^{X(a)})^{1/2}\xi_i^{X(a)}, \quad W_{ij}^{(a)}(t) \approx \Psi^{W(a)}(t)(\Delta^{W(a)})^{1/2}\xi_{ij}^{W(a)}.$$

We further decompose the eigenstructures into shared and source-specific components:

$$\begin{aligned} X_i^{(a)}(t) &\approx \Phi^X(t)U^{X(a)}(\Lambda^X)^{1/2}(D^{X(a)})^{1/2}\xi_i^{X(a)}, \\ W_{ij}^{(a)}(t) &\approx \Phi^W(t)U^{W(a)}(\Lambda^W)^{1/2}(D^{W(a)})^{1/2}\xi_{ij}^{W(a)}, \end{aligned} \tag{4}$$

where $\Phi^X(t) = (\phi_1^X(t), \ldots, \phi_{K_X}^X(t))$ and $\Phi^W(t) = (\phi_1^W(t), \ldots, \phi_{K_W}^W(t))$ are common eigenfunctions shared across sources, satisfying $\int(\Phi^X(t))^T\Phi^X(t)dt = I_{K_X}$ and $\int(\Phi^W(t))^T\Phi^W(t)dt = I_{K_W}$, and Λ^X and Λ^W are diagonal matrices containing common eigenvalues. The matrices $U^{X(a)}$ and $U^{W(a)}$ are source-specific orthogonal rotations, and the diagonal matrices $D^{X(a)}$ and $D^{W(a)}$ represent source-specific scale adjustments, typically arising when the same or similar type of data is collected by different device models or settings.

The set of unknown parameters is estimated by minimizing the objective function described in Section 2.2. To remove source-specific effects while preserving shared biological signals, we construct two forms of harmonized data. The first version, referred to as INTACT, retains residual noise to better approximate the variability of the original data:

$$\hat{Y}_{ij}^{\text{intact}(a)}(t) = c_0\Big(\hat{\alpha}(t)+(Z_{ij}^{(a)})^T\hat{\beta}(t)+c_X^{(a)}\hat{\Phi}^X(t)(\hat{\Lambda}^X)^{1/2}\hat{\xi}_i^{X(a)}+c_W^{(a)}\hat{\Phi}^W(t)(\hat{\Lambda}^W)^{1/2}\hat{\xi}_{ij}^{W(a)}+c_\epsilon^{(a)}\hat{\epsilon}_{ij}^{(a)}(t)\Big).$$

The second, denoted INTACT$_0$, removes the noise term to yield a smooth denoised version:

$$\hat{Y}_{ij}^{\text{intact}_0(a)}(t) = c_0\Big(\hat{\alpha}(t) + (Z_{ij}^{(a)})^T\hat{\beta}(t) + c_X^{(a)}\hat{\Phi}^X(t)(\hat{\Lambda}^X)^{1/2}\hat{\xi}_i^{X(a)} + c_W^{(a)}\hat{\Phi}^W(t)(\hat{\Lambda}^W)^{1/2}\hat{\xi}_{ij}^{W(a)}\Big).$$

Here, c_0, $c_X^{(a)}$, $c_W^{(a)}$, and $c_\epsilon^{(a)}$ are scaling constants used to align the variance of harmonized components with that of the original data (see Section 2.2 for details). INTACT is particularly appropriate when preservation of the original data characteristics is important, such as for individual-level analyses, whereas INTACT$_0$ is better suited for population-level applications, such as PCA-based feature extraction.

2.2 Estimation and implementation of the integration method

We introduce notations used throughout. For a matrix A, let A_k denote its kth column and A_{kk} its (k,k) entry. Let e_k be the kth column of the identity matrix I_{K_X}. The notation $\|A\|_2$ refers to the spectral norm when A is a matrix, and the Euclidean (i.e., L_2) norm when A is a vector. Bold symbols represent discretized versions of functional objects evaluated on a fixed grid $\mathcal{T} = \{t_1, \ldots, t_p\}$. For example, $\boldsymbol{S}$ denotes the discretized form of function $S(s,t)$.

To extract shared latent structure across sources in longitudinal data, we separately model the X and W components. Since only $Y_{ij}^{(a)}(t)$ (and thus $\eta_{ij}^{(a)}(t)$) is observed, we estimate the sample covariances of X and W using fast multilevel functional principal component analysis (fast MFPCA) (Cui et al., 2023). We describe the estimation procedure for the X component; the method for W is analogous. Based on the approximation

$$S^{X(a)}(s,t) \approx \Phi^X(s)U^{X(a)}\Lambda^X D^{X(a)}(U^{X(a)})^T(\Phi^X(t))^T,$$

we estimate the parameters $\{\Phi^X(t), \Lambda^X, \{U^{X(a)}, D^{X(a)}, b^{(a)}\}_{a=1}^M\}$ by minimizing:

$$\begin{aligned}
\text{minimize} \ & \sum_{a=1}^{M} \int \left\| \Phi^X(s) U^{X(a)} \Lambda^X D^{X(a)} (U^{X(a)})^T (\Phi^X(t))^T - \hat{S}^{X(a)}(s,t) \right\|^2 ds\, dt \\
& + \sum_{a=1}^{M} \sum_{k=1}^{K_X} \left\{ \lambda_k \| U_k^{X(a)} - e_k \|_2^2 + \nu_k (D_{kk}^{X(a)} - b^{(a)})^2 \right\} + \omega \sum_{k=1}^{K_X} \int (\partial^2 \phi_k^X(t))^2 dt, \qquad (5) \\
\text{subject to} \ & \sum_{a=1}^{M} b^{(a)} = M, \quad b^{(a)} \geq 0.
\end{aligned}$$

Here, $\hat{S}^{X(a)}(s,t)$ denotes the smoothed sample covariance of $X_i^{(a)}(t)$; Λ^X and $\{D^{X(a)}\}$ are diagonal and nonnegative; $\Phi^X(t)$ and $\{U^{X(a)}\}$ are column orthonormal. A larger λ_k encourages alignment of $U^{X(a)}$ with the identity matrix, while a larger ν_k shrinks $D^{X(a)}$ toward the shared scale parameter $b^{(a)}$. The smoothing penalty ω enforces smoothness of the eigenfunction curves. To prioritize leading components, we impose $\lambda_1 > \cdots > \lambda_{K_X}$ and $\nu_1 > \cdots > \nu_{K_X}$. The constraint on $\{b^{(a)}\}$ ensures identifiability given the multiplicative structure of $\Lambda^X D^{X(a)}$.

In practice, we use fast MFPCA (Cui et al., 2023), an extension of the fast covariance estimation (FACE) method (Xiao et al., 2016), to obtain smooth estimates of covariance functions for multilevel functional data. Let B denote the $p \times q$ matrix of cubic B-spline basis functions evaluated on a dense grid $\mathcal{T} = \{t_1, \ldots, t_p\}$, with entries $B_{jk} = B_k(t_j)$. In FACE, the discretized smooth covariance estimate $\hat{\boldsymbol{S}}^{X(a)}$ of $S^{X(a)}(s,t)$ is expressed as $\hat{\boldsymbol{S}}^{X(a)} = \tilde{B}\Theta^{X(a)}\tilde{B}^T$, where $\tilde{B}$ is a $p \times q$ matrix transformed from B satisfying $\tilde{B}^T\tilde{B} = I_q$, and $\Theta^{X(a)}$ is a $q \times q$ positive semi-definite matrix (see Appendix A for details). This formulation motivates expanding the function $\Phi^X(t)$ in the orthogonal basis $\tilde{B}$, which allows all parameters in problem (5) to be represented in a low-dimensional space, thus enabling efficient computation.

To reduce the complexity of tuning, we set the penalty parameters λ_k and ν_k proportional to the kth diagonal element of an initial estimate Λ^{X0}: $\lambda_k = \nu_k = a_0 \Lambda_{kk}^{X0}$, where a_0 is a scaling constant selected via cross-validation (see Appendix D). For the smoothing parameter ω, we use $\omega = c/\sum_{k=1}^{K_X} \int \left(\partial^2 \phi_k^{X0}(t)\right)^2 dt$, where c is a small constant (e.g., 0.01), and $\phi_k^{X0}(t)$ denotes an initial estimate of $\phi_k^X(t)$. A detailed algorithm for solving problem (5) is provided in Appendix C.

Selecting the number of eigenfunctions is crucial in practice. For single-source data, the number of components is typically selected to capture a high proportion of the variance explained (PVE) (Di et al., 2009; Cui et al., 2023). However, in multi-source settings, the selection should also ensure high cross-source similarity of eigenfunctions. For the first k components, define PVE in source a as $\text{PVE}^{X(a)}(k) = (\pi_1^{X(a)} + \cdots + \pi_k^{X(a)})/(\pi_1^{X(a)} + \cdots + \pi_q^{X(a)})$, $a = 1, \ldots, M$, where $\pi_j^{X(a)}$ are the eigenvalues of $\Theta^{X(a)}$. Let $\text{mPVE}^X(k) = \sum_{a=1}^M \text{PVE}^{X(a)}(k)/M$, and define cross-source similarity as $\text{Sim}^X(k) = \min_{1 \leq j \leq k,\, a_1 \neq a_2} \left| \left(V_j^{X(a_1)}\right)^T V_j^{X(a_2)} \right|$ where $V_j^{X(a)}$ are the eigenfunctions of $\Theta^{X(a)}$. The number of components K_X is selected by maximizing

$$\alpha^X \text{mPVE}^X(k) + (1 - \alpha^X)\text{Sim}^X(k) \qquad (6)$$

subject to $k \geq k_{\min}^X$, $\pi_k^{X(a)}/(\pi_1^{X(a)} + \cdots + \pi_q^{X(a)}) \geq 0.01$, and $\text{Sim}^X(k) \geq 0.01$. The parameter $\alpha^X \in [0,1]$ balances explained variance and similarity. We use an analogous strategy for selecting components for the W part and set $\alpha^X = \alpha^W = 0.5$, $k_{\min}^X = k_{\min}^W = 2$. Results are robust to a range of α^X and α^W values (Appendix H).

Specifically, our integration procedure INTACT is summarized as follows.

Step 1 (Preprocessing). We vectorize the functional outcomes observed at discrete time points $\mathcal{T} = \{t_1, \ldots, t_p\}$, and define $\boldsymbol{Y}^{(a)} = \left(\boldsymbol{Y}_{11}^{(a)}, \ldots, \boldsymbol{Y}_{1r_1}^{(a)}, \ldots, \boldsymbol{Y}_{n_a1}^{(a)}, \ldots, \boldsymbol{Y}_{n_ar_{n_a}}^{(a)}\right)^T \in \mathbb{R}^{N_a \times p}$, where $\boldsymbol{Y}_{ij}^{(a)} = \left(Y_{ij}^{(a)}(t_1), \ldots, Y_{ij}^{(a)}(t_p)\right)^T$. We scale the observations as $\tilde{Y}_{ij}^{(a)}(t) = Y_{ij}^{(a)}(t)/\sqrt{\|\boldsymbol{S}_Y^{(a)}\|_2}$, where $\boldsymbol{S}_Y^{(a)}$ is the global sample covariance matrix of $\boldsymbol{Y}^{(a)}$. Define $\tilde{\boldsymbol{Y}}^{(a)}$ analogously to $\boldsymbol{Y}^{(a)}$. To rescale harmonized data back to the original scale, we compute the overall scaling factor $c_0 = \sqrt{\|\boldsymbol{S}_0\|_2/\|\tilde{\boldsymbol{S}}_0\|_2}$, where $\boldsymbol{S}_0$ and $\tilde{\boldsymbol{S}}_0$ are the sample covariance matrices of $\{\boldsymbol{Y}_{ij}^{(1)}, \ldots, \boldsymbol{Y}_{ij}^{(M)}\}$ and $\{\tilde{\boldsymbol{Y}}_{ij}^{(1)}, \ldots, \tilde{\boldsymbol{Y}}_{ij}^{(M)}\}$, respectively.

Step 2 (Estimate $\eta_{ij}^{(a)}(t)$ in (1)). For each t, we estimate $\alpha(t), \gamma^{(a)}(t), \beta(t)$ using REML for linear mixed effects models with subject-specific random intercept (`lmer` function in the R package `lme4`). Then we apply `gam` function in the R package `mgcv` to obtain smoothed estimates $\hat{\alpha}(t), \hat{\gamma}^{(a)}(t), \hat{\beta}(t)$. Let $\hat{\eta}_{ij}^{(a)}(t) = \tilde{Y}_{ij}^{(a)}(t) - \hat{\alpha}(t) - \hat{\gamma}^{(a)}(t) - (Z_{ij}^{(a)})^T\hat{\beta}(t)$.

Step 3 (Functional PCA of $\hat{\eta}_{ij}^{(a)}(t)$). For each source a, we apply fast MFPCA to $\hat{\eta}_{ij}^{(a)}(t)$ to obtain sample covariances for X and W parts. The corresponding subject-level and visit-level scores, $\xi_i^{X(a)}$ and $\xi_{ij}^{W(a)}$, are estimated via raw projection (see Appendix B) and denoted by $\hat{\xi}_i^{X(a)}$ and $\hat{\xi}_{ij}^{W(a)}$, respectively. The residual noise terms $\epsilon_{ij}^{(a)}(t)$ are estimated as the difference between $\hat{\eta}_{ij}^{(a)}(t)$ and its projection onto the eigenspace, and are denoted by $\hat{\epsilon}_{ij}^{(a)}(t)$.

Step 4 (Estimate common features). We solve the optimization problem (5) to obtain estimates $\hat{\Phi}^X(t)$, $\hat{\Lambda}^X$, $\hat{\Phi}^W(t)$, and $\hat{\Lambda}^W$.

Step 5 (Data integration). Let

$$\begin{aligned}
\hat{\eta}_{ij}^{\text{intact}(a)}(t) =& c_X^{(a)}\hat{\Phi}^X(t)(\hat{\Lambda}^X)^{1/2}\hat{\xi}_i^{X(a)} + c_W^{(a)}\hat{\Phi}^W(t)(\hat{\Lambda}^W)^{1/2}\hat{\xi}_{ij}^{W(a)} + c_\epsilon^{(a)}\hat{\epsilon}_{ij}^{(a)}(t),\\
\hat{\eta}_{ij}^{\text{intact}_0(a)}(t) =& c_X^{(a)}\hat{\Phi}^X(t)(\hat{\Lambda}^X)^{1/2}\hat{\xi}_i^{X(a)} + c_W^{(a)}\hat{\Phi}^W(t)(\hat{\Lambda}^W)^{1/2}\hat{\xi}_{ij}^{W(a)},
\end{aligned}$$

where the scaling factors are defined as $(c_X^{(a)})^2 = \|\sum_{a=1}^M \hat{\boldsymbol{S}}^{X(a)}/M\|_2/\|\text{var}(\hat{\boldsymbol{\Phi}}^X(\hat{\Lambda}^X)^{1/2}\hat{\xi}_i^{X(a)})\|_2$, $(c_W^{(a)})^2 = \|\sum_{a=1}^M \hat{\boldsymbol{S}}^{W(a)}/M\|_2/\|\text{var}(\hat{\boldsymbol{\Phi}}^W(\hat{\Lambda}^W)^{1/2}\hat{\xi}_{ij}^{W(a)})\|_2$, $(c_\epsilon^{(a)})^2 = \|\sum_{a=1}^M \hat{\boldsymbol{S}}^{\epsilon(a)}/M\|_2/\|\hat{\boldsymbol{S}}^{\epsilon(a)}\|_2$, $\hat{\boldsymbol{S}}^{X(a)}, \hat{\boldsymbol{S}}^{W(a)}, \hat{\boldsymbol{S}}^{\epsilon(a)}$ are the discretized versions of the sample covariance functions, and $\hat{\boldsymbol{\Phi}}^X$ and $\hat{\boldsymbol{\Phi}}^W$ denote the discretized eigenfunctions. The final harmonized data are then constructed as

$$\begin{aligned}
&\text{(INTACT)} \quad \hat{Y}_{ij}^{\text{intact}(a)}(t) = c_0\left(\hat{\alpha}(t) + (Z_{ij}^{(a)})^T\hat{\beta}(t) + \hat{\eta}_{ij}^{\text{intact}(a)}(t)\right),\\
&\text{(INTACT}_0\text{)} \quad \hat{Y}_{ij}^{\text{intact}_0(a)}(t) = c_0\left(\hat{\alpha}(t) + (Z_{ij}^{(a)})^T\hat{\beta}(t) + \hat{\eta}_{ij}^{\text{intact}_0(a)}(t)\right),
\end{aligned}$$

where c_0 is the overall scaling factor defined in Step 1.

2.3 Incomplete data

Physical activity data often contain missing values, which can affect the first three steps in Section 2.2. To ensure data quality, we retain only days with sufficient wear time (more than 10 hours; see Sections 4 and 5). Assuming data are missing at random, we handle incompleteness as follows. In Step 1, we apply the `softImpute` function from the R package `softImpute` (Hastie et al., 2015) to $\boldsymbol{Y}$, yielding a low-rank approximation via nuclear-norm regularization, and use the squared leading singular value (scaled by sample size) as the spectral norm estimate of $\boldsymbol{S}_Y$. In Step 2, the `lmer` function used at each time point t automatically excludes missing observations. In Step 3, fast MFPCA accommodates missing data via the FACE procedure (Xiao et al., 2016), which iteratively smooths trajectories and imputes missing

values using best linear unbiased prediction. Simulation studies demonstrate that the proposed method remains robust to missingness; see Appendix F.3.

3 Simulations

We generate data according to the model

$$Y_{ij}^{(a)}(t) = \alpha(t) + \gamma^{(a)}(t) + Z_i^{(a)}\beta_1(t) + \eta_{ij}^{(a)}(t), \quad a = 1, \ldots, M, \quad i = 1, \ldots, n_a, \quad j = 1, \ldots, r_i,$$

where $\eta_{ij}^{(a)}(t) = \Phi^X(t)U^{X(a)}(\Lambda^X)^{1/2}(D^{X(a)})^{1/2}\xi_i^{X(a)} + \Phi^W(t)U^{W(a)}(\Lambda^W)^{1/2}(D^{W(a)})^{1/2}\xi_{ij}^{W(a)} + \epsilon_{ij}^{(a)}(t)$. Functional observations are recorded at $t \in \{1, \ldots, 144\}$. Let $\mathcal{T}_i^{(a)}$ and $C_i^{(a)}$ denote the true and censoring times, respectively, with the observed time defined as $\widetilde{\mathcal{T}}_i^{(a)} = \min(\mathcal{T}_i^{(a)}, C_i^{(a)})$. The subject-specific mean curve is $q_i^{(a)}(t) = \alpha(t) + Z_i^{(a)}\beta_1(t) + \Phi^X(t)(\Lambda^X)^{1/2}\xi_i^{X(a)} + \Phi^W(t)(\Lambda^W)^{1/2}\bar{\xi}_i^{W(a)} + \bar{\epsilon}_i^{(a)}(t)$, where $\bar{\xi}_i^{W(a)} = r_i^{-1}\sum_j \xi_{ij}^{W(a)}$ and $\bar{\epsilon}_i^{(a)}(t) = r_i^{-1}\sum_j \epsilon_{ij}^{(a)}(t)$. The survival time follows a Cox model $P(\mathcal{T}_i^{(a)} > \tau) = \exp\left\{-\tau \exp\left(\beta_2^T Q(q_i^{(a)}(t))\right)\right\}$, where $Q(q_i^{(a)}(t))$ denotes a functional summary of the curve.

We set $M = 2$, $\gamma^{(a)}(t) = 0$, $n_a = 200$, $C_i^{(a)} = 5$, and $r_i \in \{3, 4, 5, 6, 7\}$. The common eigenfunctions are defined as $\Phi^X(t) = (\phi_1^X(t), \phi_2^X(t), \phi_3^X(t))$ and $\Phi^W(t) = (\phi_1^W(t), \phi_2^W(t), \phi_3^W(t))$. The shapes of $\alpha(t)$, $\beta_1(t)$, the common eigenfunctions $\Phi^X(t)$ and $\Phi^W(t)$, as well as the source-specific eigenfunctions, are presented in Figure 1A. Let $R_x(\theta),\ R_y(\theta),\ R_z(\theta)$ denote the basic 3D rotation matrices (see Appendix F.1 for detailed definitions). We set

$$\begin{aligned}
&U^{X(1)} = R_z(\pi/6)R_y(0.35)R_x(0.2), \quad U^{X(2)} = R_x(-\pi/6)R_y(-0.35)R_z(0.2),\\
&\Lambda^X = \text{diag}(1.8, 1.2, 0.5), \quad D^{X(1)} = \text{diag}(0.9, 0.8, 0.8), \quad D^{X(2)} = \text{diag}(1.1, 1.2, 1.2),\\
&U^{W(1)} = R_z(0.35)R_y(\pi/6)R_x(0.3), \quad U^{W(2)} = R_x(-0.3)R_y(-\pi/6)R_z(-0.15),\\
&\Lambda^W = \text{diag}(1, 0.7, 0.5), \quad D^{W(1)} = \text{diag}(1.2, 1.1, 1.1), \quad D^{W(2)} = \text{diag}(0.8, 0.85, 0.85),
\end{aligned}$$

and heatmaps of source-specific rotations and scalings are provided in Appendix F.2. Let $Z_i^{(a)} \overset{iid}{\sim} N(1, 0.5^2)$, $\xi_i^{X(a)} \overset{iid}{\sim} N(0, I_3)$, $\xi_{ij}^{W(a)} \overset{iid}{\sim} N(0, I_3)$, and $\epsilon_{ij}^{(a)}(t) \overset{iid}{\sim} N(0, 0.2^2)$, $a = 1, 2$. We consider three choices for $Q(\cdot)$ in the Cox model: (C1) $Q(q_i^{(a)}(t)) = \int q_i^{(a)}(t)dt$ with $\beta_2 = 0.6$; (C2) $Q(q_i^{(a)}(t)) = \int(\exp(q_i^{(a)}(t)) - 1)dt$ with $\beta_2 = 0.2$; (C3) $Q(q_i^{(a)}(t)) = \left(\int q_i^{(a)}(t)dt, \int(\exp(q_i^{(a)}(t)) - 1)dt\right)^T$ with $\beta_2 = (0.6, 0.2)^T$. To evaluate robustness to incomplete data, we also consider a scenario with 30% of time points missing at random. The results are similar to those under complete data and are presented in Appendix F.3.

To ensure comparability, we scale the original data $Y_{ij}^{(a)}(t)$ by the square root of the spectral norm of its sample covariance matrix (Step 1 in Section 2.2). The resulting scaled data $\tilde{Y}_{ij}^{(a)}(t)$ serve as the baseline and are referred to as the raw data for fair comparison. We compare INTACT to three alternative harmonization methods applied to the raw data: simple standardization, quantile normalization (Amaratunga and Cabrera, 2001; Bolstad et al., 2003), and longitudinal ComBat (Beer et al., 2020). The first two approaches are implemented pointwise and do not account for the temporal correlation and within-subject dependence in the data (see Appendix E for details). In contrast, INTACT explicitly models both longitudinal structure and temporal patterns, making it more appropriate for longitudinal time-varying data integration.

Table 1: Simulation results for complete data, including mean computation time (Time, seconds); mean variance discrepancies (MV^X, DV^X, MV^W, DV^W); empirical power for detecting source effects (Source-effect); and mean squared errors (C1–C3 MSE) and empirical power (C1–C3 Power) for β_2 across three Cox models.

	raw	standard	quantile	L-ComBat	INTACT	INTACT_0
Time (seconds)	0.006	1.950	2.022	2.228	2.102	2.102
MV^X	0.696	0.569	0.572	0.752	0.402	0.335
DV^X	2.151	1.628	1.572	2.423	0.052	0.058
MV^W	0.225	0.228	0.341	0.248	0.188	0.159
DV^W	0.570	0.577	0.479	0.583	0.004	0.005
Source-effect	0.665	0.680	0.995	0.620	0.135	0.025
C1 MSE	0.164	0.224	0.249	0.219	0.161	0.160
C2 MSE	0.004	0.003	0.003	0.003	0.003	0.003
C3 MSE	11.367	10.304	11.514	10.733	10.249	10.212
C1 Power	0.233	0.275	0.283	0.292	0.308	0.308
C2 Power	0.600	0.850	0.842	0.867	0.908	0.900
C3 Power	0.767	0.975	0.975	0.975	1.000	1.000

Under the simulation setting described above, we generate 200 datasets. Results for the complete data scenario are presented in Figure 1 and Table 1. The first row of Table 1 reports the runtime of each method, measured on a MacBook Pro (13-inch, 2022) with an Apple M2 chip and 24 GB of memory. All integration methods exhibit comparable runtimes. For INTACT, the main computational cost arises from fixed-effects estimation (Step 2 in Section 2.2), which takes 1.938 seconds.

The number of eigenfunctions is selected using the criterion (6). As shown in Figure 1B, the estimated numbers $\hat{K}_X$ and $\hat{K}_W$ do not always match the true value of 3, which is expected due to basis function approximation and the presence of noise $\epsilon_{ij}^{(a)}(t)$. Figures 1C and 1D demonstrate that the estimated curves for $\beta_1(t)$, $\Phi^X(t)$, and $\Phi^W(t)$ closely align with the truth. To quantify the estimation accuracy of the covariance structures, we compute:

$$\text{MV}^\iota = \frac{1}{M}\sum_{a=1}^{M} d_F^2\left(\hat{S}^{\iota(a)}, \Phi^\iota(t)\Lambda^\iota\Phi^\iota(t)^T\right), \quad \text{DV}^\iota = \frac{1}{M(M-1)}\sum_{a_1\neq a_2} d_F^2\left(\hat{S}^{\iota(a_1)}, \hat{S}^{\iota(a_2)}\right),$$

where $\iota \in \{X, W\}$, d_F is the Frobenius norm, and $\hat{S}^{\iota(a)}$ denotes the sample covariance from source a, estimated via fast MFPCA. Rows 2-5 of Table 1 show the mean values of MV^ι and DV^ι. We see that both INTACT and INTACT_0 achieve substantially lower mean covariance discrepancies in terms of both MV^ι and DV^ι. For the survival analysis, we estimate β_2 using the `coxph` function from the `survival` R package and evaluate performance via mean squared error (MSE) and empirical power for testing $\beta_2 = 0$ at the 0.05 significance level. As shown in the last six rows of Table 1, INTACT and INTACT_0 yield slightly lower MSE and higher power compared to alternative methods.

We next evaluate harmonization effectiveness across sources. For each simulated dataset, one observation (i.e., one day) is randomly selected per subject to form a resampled dataset of independent observations. To remove potential confounding due to covariate effects, we subtract the estimated fixed effects $\hat{\alpha}(t) + Z_i^{(a)}\hat{\beta}_1(t)$ from the harmonized data. To assess distributional similarity, we apply a generalized graph-based two-sample test (Chen and Friedman, 2017) using the minimum spanning tree as the similarity graph. This test evaluates whether the distributions of the harmonized data differ across sources. The sixth row of Table 1 reports empirical power at the 0.05 significance level. Lower power

for INTACT and INTACT$_0$ indicates more effective harmonization. We further assess source separability using a random forest–based classification approach (Breiman, 2001), as also adopted by Chen et al. (2022). Each dataset is randomly split into 50% training and 50% testing sets. A random forest classifier (R package `randomForest`, default settings) is trained to predict source labels and evaluated via the area under the ROC curve (AUC) and the adjusted Rand index (ARI). Lower AUC and ARI values indicate reduced source distinguishability and thus better harmonization. As shown in Figure 1E, both INTACT and INTACT$_0$, along with simple standardization and longitudinal ComBat, substantially reduce source separability compared to unharmonized and quantile-normalized data.

Finally, we evaluate the ability to detect biological signals by testing the null hypothesis $H_0 : \beta_1(t) = 0$ at each time point t. We employ the empirical likelihood ratio test of Wang et al. (2010) and report the empirical power at the 0.05 significance level in Figure 1F. The proposed harmonized data, $\hat{Y}_{ij}^{\text{INTACT}(a)}(t)$ and $\hat{Y}_{ij}^{\text{INTACT}_0(a)}(t)$, achieve higher power and detect more significant effects across nearly the entire time domain.

4 Integration of NHANES 2003-2004 and 2013-2014 data

We evaluate the proposed method using physical activity data from NHANES 2003–2004 and NHANES 2013–2014. The former employed a hip-worn Actigraph AM-7164 accelerometer, while the latter used a wrist-worn Actigraph GT3X+. The 2003–2004 cycle includes 7,176 participants aged 6 years and older, and the 2013–2014 cycle includes 7,776 participants aged 3 years and older, with seven consecutive days of monitoring in both studies. A major challenge in integrating the two NHANES datasets is the discrepancy in measurement units. The 2003–2004 data report activity counts (mean = 222.2), while the 2013–2014 data use the MIMS metric (mean = 8.863), which captures intensity on a smaller scale. Although MIMS is device-agnostic and supported by open-source software, it requires tri-axial input and is thus inapplicable to the uni-axial accelerometer used in 2003–2004.

In this study, we aim to integrate longitudinal minute-level physical activity data from two NHANES accelerometer devices to enhance detection of biologically meaningful signals. We focus on participants aged 50 years and older, with activity data linked to covariates $Z_{ij}^{(a)} = (\text{gender}, \text{age}, \text{income}, \text{BMI})^T$, where gender is coded as 0 for female and 1 for male. To ensure data quality, we exclude eight participants from the NHANES 2003–2004 cohort with extreme activity values and retain only days with at least 10 hours of estimated wear time. Daily activity is summarized into 1,440-dimensional vectors, corresponding to minute-level measurements across a 24-hour period. The NHANES physical activity data are further linked to 2019 mortality outcomes using the Public-use Linked Mortality Files.

After preprocessing, the final dataset includes 1,870 participants (11,125 observation days) from NHANES 2003–2004 and 1,986 participants (12,961 observation days) from NHANES 2013–2014. We apply a log transformation to activity intensities and evaluate multiple harmonization methods, including INTACT, INTACT$_0$, simple standardization, quantile normalization, and longitudinal ComBat. Figure 2A displays the mean activity curves for the two datasets, showing similar temporal patterns but clear magnitude differences. As illustrated in Figure 3B, the leading eigenfunctions of the two datasets are closely aligned, supporting the suitability of the proposed model for harmonization.

4.1 Computation time and estimation of model parameters

The first row of Table 2(a) reports computation times for all methods under the same setup as the simulations, showing comparable runtimes. For INTACT, the main cost is fixed-effect estimation (Step 2, Section 2.2), requiring 145.317 seconds.

The first row of Figure 3A presents the estimated common mean curve and source-specific shifts. As expected, activity levels are lower during nighttime and higher during the day. The second row shows the estimated fixed-effect functions for gender, age, income, and BMI. Males exhibit more nighttime and less daytime activity relative to females; older age is associated with reduced activity across the day. Higher income is linked to increased daytime and reduced nighttime activity, while higher BMI shows the opposite trend.

The number of eigenfunctions is selected via criterion (6) with weights $\alpha^X = \alpha^W = 0.5$, resulting in three components for the X part and two for the W part. Results are robust to alternative weight choices (see Appendix H). The solid red curves in Figure 3B display the estimated eigenfunctions. The first eigenfunction for the X part indicates that participants with positive scores tend to exhibit lower nighttime activity and higher daytime activity. The second eigenfunction is negative during early morning hours (4am–10am), suggesting that positive scores are associated with reduced morning activity and elevated activity during the remainder of the day. The third eigenfunction reflects lower activity during typical working hours (9am–6pm) and higher activity outside this period. The eigenfunctions for the W part capture within-subject (day-to-day) variability. Days with positive scores on the first eigenfunction show elevated activity from 4am to midnight relative to the individual's average. For the second eigenfunction, positive scores correspond to decreased morning activity and increased activity from 10am through midnight. The estimated source-specific rotations and scale adjustments are provided in Appendix H.

4.2 Evaluation of source effects

Figure 2B displays the overall mean curves for the raw and harmonized data, showing that all methods effectively reduce mean differences between sources. Sample covariances are estimated using fast MFPCA, and the corresponding heatmaps are shown in Appendix H. Figure 2C presents the differences between the source-specific covariance matrices. INTACT and INTACT$_0$ yield nearly identical covariance structures across sources, whereas the raw and other harmonized data exhibit greater discrepancies.

To further evaluate harmonization, we perform a resampling procedure by randomly selecting 150 participants per source and drawing one day per participant to construct a resampled dataset of independent observations. To mitigate confounding, we subtract estimated fixed effects $\hat{\alpha}(t) + (Z_{ij}^{(a)})^T \hat{\beta}(t)$ from the harmonized data. This process is repeated 200 times. Each resampled dataset is evaluated using the generalized graph-based test and a random forest classifier, as in the simulation study. The second row of Table 2(a) reports the empirical power for rejecting the null hypothesis of equal distributions at the 0.05 significance level. INTACT, INTACT$_0$, and simple standardization achieve the lowest power, indicating superior harmonization. Figure 3C shows boxplots of AUC and ARI from the random forest classification. INTACT, INTACT$_0$, and quantile normalization result in substantially lower values, reflecting reduced source separability.

4.3 Age effects in physical activity

Previous research has shown that age influences circadian activity patterns (Xiao et al., 2015; Wrobel et al., 2021). To assess whether the harmonized data enhance detection of age-related associations, we test the null hypothesis of no age effect on activity intensity. A resampling procedure is conducted by randomly selecting 150 participants from each source and applying a pointwise empirical likelihood ratio test (Wang et al., 2010) within a linear mixed-effects model framework. This procedure is repeated 200 times to estimate the empirical power for detecting age effects. Figure 3D presents the results. Using data from NHANES 2003–2004 or 2013–2014 alone yields empirical power near zero, as shown by the dark gray dashed and dotted lines, indicating insufficient signal. In contrast, combining data

from both sources—whether raw or harmonized—enables detection of significant age effects. Among all approaches, INTACT and INTACT_0 consistently achieve the highest power across the time domain, indicating superior ability to achieve biologically meaningful signals.

4.4 Association between mortality and physical activity

Recent research has increasingly explored the relationship between mortality and physical activity patterns (Cui et al., 2021). To examine this association, we fit a Cox proportional hazards model to hourly average log-transformed physical activity intensities, adjusting for gender, age, income, and BMI, using a 5-year censoring period. The model is implemented via the `coxph()` function in the R package `survival`. For participants with multiple days of measurements, we average the log-transformed intensities across days prior to model fitting.

We first fit separate Cox models to activity intensities at each hour of the day. Figure 3E displays the estimated hazard ratios and Bonferroni-corrected 99.79%(=1-0.05/24) confidence intervals for each hour. All methods consistently show significant protective effects of daytime physical activity (approximately 8am to 10pm), with hazard ratios below 1, indicating that higher activity during these hours is associated with reduced mortality risk. From 12am to 6am, hazard ratios for INTACT and INTACT_0 exceed 1, while those from other methods remain around or above 1 but are not statistically significant. Notably, only the proposed methods (INTACT and INTACT_0) detect significant associations between 2am and 5am, suggesting that higher early-morning activity may be detrimental to longevity. Similar results based on minute-level data are provided in Appendix H, demonstrating the robustness of our findings. To further explore the predictive association between overall activity patterns and mortality, we summarize each individual's activity profile by averaging intensities over 2-hour intervals, resulting in a 12-dimensional feature vector. We then fit a Cox model using this activity vector and the same set of covariates. Columns 2-4 of Table 2(b) reports the mean 10-fold cross-validated Harrell's C-index. INTACT and INTACT_0 achieve the highest C-index values among all methods, indicating superior predictive performance in assessing mortality risk.

4.5 Downstream analysis of summary statistics following data integration

Since many wearable device studies derive subject-level activity markers for downstream analyses, we evaluate whether epoch-level harmonization improves the consistency of summary statistics between NHANES 2003–2004 and NHANES 2013–2014. We randomly select 1,000 participants from NHANES 2003–2004 and match them to 1,000 participants from NHANES 2013–2014 on gender, age, income, and BMI using the `pairmatch()` function from the R package `optmatch` (Hansen, 2007), with distances computed via `match_on()` using logit propensity scores. For each participant, one day is randomly chosen from the repeated measurements to compute six common summary statistics: total activity counts (TAC), total log-transformed activity counts (TLAC), mean activity during the most active 10-hour period (M10), timing of M10 (M10t), mean activity during the least active 5-hour period (L5), and timing of L5 (L5t). We then calculate the absolute differences in means and variances of these statistics between the two sources.

This resampling and evaluation procedure is repeated 200 times. Figure 3F presents boxplots ($\log_{10}$ scale) of mean and variance differences for each summary statistic. The raw data exhibit substantial inconsistencies across sources. Simple standardization and longitudinal ComBat show similar performance, reducing mean differences in TLAC, L5, and L5t but increasing those in TAC and M10, along with larger variance differences for all statistics except TLAC. Quantile normalization improves mean differences for most statistics and reduces the variance difference in L5, but leaves other variance measures largely unchanged. In contrast, both INTACT and INTACT_0 achieve markedly lower mean and

Table 2: Integration results for real-data applications. Panels (a) and (b) correspond to integration of NHANES 2003–2004 and NHANES 2013–2014 accelerometer data. Panel (c) corresponds to integration of NHANES and ExtraSensory accelerometer data.

(a) Computation time (Time, seconds) and empirical power for source-effect detection (Source-effect) when integrating NHANES 2003–2004 and NHANES 2013–2014 accelerometer data.

	raw	standard	quantile	L-ComBat	INTACT	INTACT_0
Time (seconds)	0.498	146.738	156.077	144.670	160.310	160.310
Source-effect	0.915	0.690	1.000	0.785	0.550	0.090

(b) Average 10-fold cross-validated Harrell's C-index for Cox models using either a 12-dimensional activity vector with gender, age, income, and BMI (columns 2–4) or daily summary statistics (TAC, TLAC, M10, M10t, L5, L5t) with gender, age, income, and BMI (columns 5–7).

	source 1	source 2	all	source 1	source 2	all
raw	0.790	0.773	0.777	0.780	0.755	0.754
standard	0.790	0.774	0.782	0.779	0.756	0.758
quantile	0.789	0.774	0.783	0.781	0.756	0.759
L-ComBat	0.790	0.773	0.782	0.779	0.756	0.758
INTACT	0.794	0.778	0.786	0.781	0.758	0.760
INTACT_0	0.795	0.778	0.786	0.784	0.758	0.760

(c) Computation time (Time, seconds) for integrating NHANES and ExtraSensory accelerometer data, and empirical power for detecting TAC and TLAC differences between two sleep onset groups.

	raw	standard	quantile	L-ComBat	INTACT	INTACT_0
Time (seconds)	0.504	6.735	10.436	48.336	12.048	12.048
TAC	0.096	0.196	0.518	0.161	0.554	0.554
TLAC	0.161	0.089	0.429	0.089	0.554	0.554

variance differences. INTACT matches or improves upon the raw data in mean differences while attaining nearly the second-lowest variance differences across all measures. The denoised INTACT_0 consistently yields the smallest mean and variance differences, indicating superior harmonization.

We further assess the association between these summary statistics and mortality. For each matched dataset, we fit a Cox model using the activity summaries (TAC, TLAC, M10, M10t, L5, L5t) and covariates (gender, age, income, BMI), and compute the mean 10-fold cross-validated Harrell's C-index. Columns 5-7 of Table 2(b) presents the average C-index across 200 matched datasets. Across all scenarios—using source 1 (NHANES 2003–2004), source 2 (NHANES 2013–2014), or the combined data—INTACT and INTACT_0 consistently yield slightly higher predictive performance than competing methods.

5 Integration of NHANES and ExtraSensory data

To evaluate the performance of the proposed method in integrating datasets that include commercial devices measuring different physical quantities, we apply INTACT (and INTACT_0) to the NHANES 2013–2014 physical activity data and the ExtraSensory dataset. Both datasets provide minute-level activity measurements and include sleep annotations. The ExtraSensory dataset contains data from 60 participants aged 18–42 years with BMIs between 18 and 32, who wore their own smartphones and smartwatches for 2.9–28.1 days (Vaizman et al., 2017). It includes minute-level measurements from multiple sensors, such as smartphone gyroscopes and smartwatch accelerometers.

We select NHANES 2013–2014 participants aged 18–42 years with BMIs of 18-32 to match the ExtraSensory cohort. Here individual demographic covariates are not adjusted in functional regression models as individual-specific variables were not available from ExtraSensory. NHANES activity intensities and ExtraSensory gyroscope and accelerometer measurements are each summarized into 1,440-dimensional daily vectors. Figure 4A presents the mean curves for these three data types. Despite differences in magnitudes, all exhibit a similar pattern of elevated daytime activity and lower nighttime activity.

To ensure data quality, we retain only days with at least 10 hours of estimated wear time. After preprocessing, the dataset includes 1,240 participants (7,656 observation days) from NHANES 2013–2014, 50 participants (184 days) from the ExtraSensory gyroscope data, and 38 participants (108 days) from the ExtraSensory accelerometer data. All activity measures are log-transformed prior to harmonization. Subsequent analyses focus on integrating NHANES and ExtraSensory accelerometer data; results for the NHANES accelerometer and ExtraSensory gyroscope integration are similar and reported in Appendix I.

5.1 Parameter estimation and source effect evaluation

The first row of Table 2(c) summarizes computation times, showing that longitudinal ComBat is the most time-consuming method, whereas all others complete within 13 seconds. Figure 4B presents the estimated common mean curve and source-specific mean shift functions, revealing lower nighttime and higher daytime activity. Using the selection criterion in (6), we retain 2 eigenfunctions for the X part and 9 for the W part. Figure 4C presents the two estimated eigenfunctions for the X part and the first three for the W part. For the X part, the first eigenfunction reflects elevated daytime activity, while the second is positive throughout the day except from 4 am to 9 am, indicating reduced early-morning activity. The W part captures within-subject variation: the first eigenfunction corresponds to increased activity from early morning to afternoon; the second indicates reduced activity from midnight to 8 am and elevated activity otherwise; and the third reflects lower working-hour activity relative to the individual's average pattern.

We next assess source effects in the harmonized data. Figure 4D displays the overall mean curves, showing that harmonization markedly reduces mean differences between sources compared to the raw data. We estimate sample covariances using fast MFPCA, and Figure 4E presents heatmaps of covariance differences. INTACT and INTACT$_0$ produce minimal covariance discrepancies, whereas raw data and other methods exhibit larger differences.

5.2 Association between sleep onset time and physical activity

Previous studies indicate that sleep onset time influences next-day physical activity patterns (Master et al., 2019; Leota et al., 2025). To assess whether harmonization improves detection of such associations, we test the null hypothesis of no sleep onset effect on daily activity intensity. Minute-level sleep annotations classify each day as before or after midnight sleep onset. One day per subject is randomly selected, yielding 788 days in Group 1 (before 12 am) and 484 days in Group 2 (after 12 am). Next-day activity intensities are summarized by TAC and TLAC, and compared between groups using the Wilcoxon test (`wilcox.test()` in `stats`). This procedure is repeated 200 times, and the second and third rows of Table 2(c) report empirical power at the 0.05 level. INTACT and INTACT$_0$ show substantially greater power than others, highlighting superior sensitivity to sleep onset associations.

6 Discussion

In this paper, we propose INTACT, a data integration method that accounts for both longitudinal and within-day correlations in physical activity data. We demonstrate its effectiveness, along with that of its denoised variant INTACT$_0$, through simulations and applications to data collected with different generations of wearable devices. Our model assumes a common eigenspace across sources. This assumption can be assessed using prior knowledge (e.g., whether sources measure the same or related biosignals) and by examining whether mean curves and source-specific eigenfunctions exhibit comparable patterns. When this assumption is violated, numerical results in Appendices G and I show that INTACT performs well when source-specific eigenspaces are reasonably aligned. When eigenfunctions differ substantially—as in multi-source data capturing distinct underlying biosignals—INTACT could still sufficiently remove source effects. However, preservation of biosignals may be compromised (Appendix G). In addition, for simultaneous multi-sensor data, while it would be valuable to study cross-sensor relationships or joint features, this problem lies beyond the current scope, as our harmonization framework is primarily designed to increase effective sample size. Investigation of this scenario is left for future work.

For large-scale datasets with many subjects and dense time points, computational cost may become a concern. Fixed-effect estimation (Step 2) is the main bottleneck, requiring $O(g_0 Np)$ operations, where g_0 is the iteration count (typically a few dozen), $N = \sum_{a=1}^{M} N_a$ is the total number of observations, and p is the number of time points. This step can be parallelized across time points to improve efficiency. Covariance estimation (Step 3) also scales with data size, as fast MFPCA requires $O(Np)$ operations. Nevertheless, Cui et al. (2023) report that it completes in about 10 seconds for $Np = 5 \times 10^5$ on a regular laptop. Since the functional data objects are expanded using the orthogonal basis $\tilde{B}$ from fast MFPCA, the computational cost of the core step (Step 4) will remain modest with large data size (see Appendix C for implementation details).

A future research direction is to extend this method to data from different device brands, studies, or populations (e.g., diseased vs. healthy, children vs. adults). Moreover, the current model assumes a common time resolution (e.g., minute-level). While we can often align and interpolate data collected at different resolutions (e.g., UK Biobank data at 5-second intervals vs. minute intervals), the impact of

such preprocessing has not been systematically studied. Assessing this remains an important avenue for future work.

The INTACT framework is potentially applicable to other longitudinal time-varying data that require harmonization, provided preprocessing and model specification are guided by prior knowledge. The R package `intactPA` is available at `https://github.com/jingru-zhang/intactPA`.

Acknowledgements

We thank the Editor, Associate Editor, and reviewers for their insightful suggestions, which have greatly improved the quality of this paper. Dr. Zhang was supported in part by NSFC grant 12401388 and Shanghai Pujiang Program grant 23PJ1401100. Dr. Li was supported in part by NIH grant GM129781.

Conflict of interest

None declared.

Data availability

The National Health and Nutrition Examination Survey (NHANES) data used in this study are publicly available from the NHANES website (`https://wwwn.cdc.gov/nchs/nhanes/`), with a pre-processed, analysis-ready version accessible at `https://functionaldataanalysis.org/dataset_nhanes.html` (Crainiceanu et al., 2024). The Public-use Linked Mortality Files are available at `https://www.cdc.gov/nchs/data-linkage/mortality-public.htm`. The ExtraSensory data are available at `http://extrasensory.ucsd.edu`.

APPENDIX

A Covariance estimation

We adopt the fast multilevel functional principal component analysis (fast MFPCA) (Cui et al., 2023) to obtain smooth covariance estimates. Fast MFPCA decomposes the sample covariance into two levels—between-subject (X part) and within-subject (W part)—and applies the fast covariance estimation (FACE) method (Xiao et al., 2016) separately to each. Below, we review the estimation procedure for the X part; the W part is handled analogously.

In FACE, P-splines are used to construct smooth covariance estimates. Let B be the $p \times q$ matrix of B-spline basis functions evaluated at a fine knot grid $\mathcal{T} = \{t_1, \ldots, t_p\}$, with entries $B_{jk} = B_k(t_j)$. Let

$$D_2 = \begin{pmatrix} 1 & -2 & 1 & & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{pmatrix}$$

be the second-order difference operator, and $D_{2,q}$ its $(q-2) \times q$ version. The symmetric penalty matrix is $P = D_{2,q}^T D_{2,q}$ of size $q \times q$. FACE begins with the eigen decomposition $(B^T B)^{-1/2} P (B^T B)^{-1/2} = U_B \operatorname{diag}(s_1, \ldots, s_q)\, U_B^T$, and defines $\tilde{B} = B(B^T B)^{-1/2} U_B$, a $p \times q$ matrix satisfying $\tilde{B}^T \tilde{B} = I_q$. Let $\hat{\mathbf{S}}^{\mathbf{X(a)}}$ denote the discrete version of $\hat{S}^{X(a)}(s,t)$. FACE then expresses $\hat{\mathbf{S}}^{\mathbf{X(a)}} = \tilde{B}\, \Theta^{X(a)}\, \tilde{B}^T$, where $\Theta^{X(a)}$ is a $q \times q$ positive semi-definite matrix.

B Principal score estimation

We apply fast MFPCA (Cui et al., 2023) to estimate the eigenfunctions $\hat{\Psi}^{X(a)}(t)$, $\hat{\Psi}^{W(a)}(t)$ and eigenvalues $\hat{\Delta}^{X(a)}$, $\hat{\Delta}^{W(a)}$. Principal scores for the X and W parts are obtained via raw projection in the model

$$\begin{aligned}\eta_{ij}^{(a)}(t) &= X_i^{(a)}(t) + W_{ij}^{(a)}(t) + \epsilon_{ij}^{(a)}(t) \\ &= \Psi^{X(a)}(t)\big(\Delta^{X(a)}\big)^{1/2}\xi_i^{X(a)} + \Psi^{W(a)}(t)\big(\Delta^{W(a)}\big)^{1/2}\xi_{ij}^{W(a)} + \epsilon_{ij}^{(a)}(t),\end{aligned} \tag{B.1}$$

where $\xi_i^{X(a)}$ and $\xi_{ij}^{W(a)}$ are mutually independent with mean zero and identity covariance. From (B.1),

$$E_{\xi_{ij}^{W(a)}}\left[\eta_{ij}^{(a)}(t) \mid \xi_i^{X(a)}\right] = \Psi^{X(a)}(t)\big(\Delta^{X(a)}\big)^{1/2}\xi_i^{X(a)},$$

which leads to the estimator

$$\hat{\xi}_i^{X(a)} = \big(\hat{\Delta}^{X(a)}\big)^{-1/2}\big(\hat{\Psi}^{X(a)}(t)\big)^T\left(\frac{1}{r_i}\sum_{j=1}^{r_i}\hat{\eta}_{ij}^{(a)}(t)\right). \tag{B.2}$$

Substituting (B.2) into (B.1), the W part scores are estimated by

$$\hat{\xi}_{ij}^{W(a)} = \big(\hat{\Delta}^{W(a)}\big)^{-1/2}\big(\hat{\Psi}^{W(a)}(t)\big)^T\left[\hat{\eta}_{ij}^{(a)}(t) - \hat{\Psi}^{X(a)}(t)\big(\hat{\Delta}^{X(a)}\big)^{1/2}\hat{\xi}_i^{X(a)}\right].$$

C Common feature estimation

Let $\Theta^{X(a)} = V^{X(a)}\Pi^{X(a)}(V^{X(a)})^T$ be the eigendecomposition of $\Theta^{X(a)}$, where $\Pi^{X(a)} = \mathrm{diag}(\pi_1^{X(a)}, \ldots, \pi_q^{X(a)})$ and $V^{X(a)} = (V_1^{X(a)}, \ldots, V_q^{X(a)})$. The eigenvectors of $\hat{\mathbf{S}}^{X(a)}$ are then given by $\tilde{B}V_k^{X(a)}$ for $k = 1, \ldots, q$. This motivates expanding $\Phi^X(t)$ in the orthogonal basis $\tilde{B}$ when solving the optimization problem (5). Its discrete representation is assumed to be $\mathbf{\Phi^X} = \tilde{B}\,\Xi^X$, where $\Xi^X \in \mathbb{R}^{q\times K_X}$ satisfies $(\Xi^X)^T\Xi^X = I_{K_X}$. Substituting this representation into (5), we estimate the parameter set $\{\Xi^X, \Lambda^X, \{U^{X(a)}, D^{X(a)}, b^{(a)}\}_{a=1}^M\}$ by solving

$$\begin{aligned}\text{minimize } &\sum_{a=1}^M \left\|\Theta^{X(a)} - \Xi^X U^{X(a)}\Lambda^X D^{X(a)}(U^{X(a)})^T(\Xi^X)^T\right\|_F^2 \\ &+ \sum_{a=1}^M\sum_{k=1}^{K_X}\left\{\lambda_k\|U_k^{X(a)} - e_k\|_2^2 + \nu_k\left(D_{kk}^{X(a)} - b^{(a)}\right)^2\right\} + \omega\,\mathrm{tr}\left((\Xi^X)^T B_0\Xi^X\right) \\ \text{subject to } &1^T b = M, \quad b^{(a)} \geq 0,\end{aligned} \tag{C.3}$$

where $b = (b^{(1)}, \ldots, b^{(M)})^T$, $B_0 = \tilde{B}^T D_{2,p}^T D_{2,p}\tilde{B}$, and $D_{2,p}$ is the $(p-2)\times p$ second-order difference operator.

In the following, we omit the upper index X and rewrite the objective function (C.3) as

$$\begin{aligned}&L(\Xi, \Lambda, \{D^{(a)}, U^{(a)}\}_{a=1}^M, b) \\ =&\sum_{a=1}^M\left\{-\langle\Theta^{(a)}, \Xi U^{(a)}\Lambda D^{(a)}(U^{(a)})^T\Xi^T\rangle + \frac{1}{2}\|\Lambda D^{(a)}\|_F^2\right\} + \sum_{a=1}^M\sum_{k=1}^{K_X}\left\{\frac{\lambda_k}{2}\|U_k^{(a)} - e_k\|_2^2\right. \\ &\left. + \frac{\nu_k}{2}(D_{kk}^{(a)} - b^{(a)})^2\right\} + \frac{\omega}{2}\mathrm{tr}(\Xi^T B_0\Xi).\end{aligned}$$

We define the following approximations to the objective function:

$$
\begin{aligned}
&\hat{L}_1(\Xi; \hat{\Xi}, \hat{\Lambda}, \{\hat{D}^{(a)}, \hat{U}^{(a)}\}_{a=1}^M, \hat{b}) \\
=&\sum_{a=1}^M \left\{ -\langle \Theta^{(a)}, \hat{\Xi}\hat{U}^{(a)}\hat{\Lambda}\hat{D}^{(a)}(\hat{U}^{(a)})^T\hat{\Xi}^T\rangle - 2\langle \Theta^{(a)}\hat{\Xi}\hat{U}^{(a)}\hat{\Lambda}\hat{D}^{(a)}(\hat{U}^{(a)})^T, \Xi - \hat{\Xi}\rangle + \frac{1}{2}\|\hat{\Lambda}\hat{D}^{(a)}\|_F^2 \right\} \\
&+ \sum_{a=1}^M \sum_{k=1}^{K_X} \left\{ \frac{\lambda_k}{2}\|\hat{U}_k^{(a)} - e_k\|_2^2 + \frac{\nu_k}{2}(\hat{D}_{kk}^{(a)} - \hat{b}^{(a)})^2 \right\} + \langle \omega B_0 \hat{\Xi}, \Xi - \hat{\Xi}\rangle + \frac{\omega}{2}\mathrm{tr}(\hat{\Xi}^T B_0 \hat{\Xi}),
\end{aligned}
$$

$$
\begin{aligned}
&\hat{L}_2(\{U^{(a)}\}_{a=1}^M; \hat{\Xi}, \hat{\Lambda}, \{\hat{D}^{(a)}, \hat{U}^{(a)}\}_{a=1}^M, \hat{b}) \\
=&\sum_{a=1}^M \left\{ -\langle \Theta^{(a)}, \hat{\Xi}\hat{U}^{(a)}\hat{\Lambda}\hat{D}^{(a)}(\hat{U}^{(a)})^T\hat{\Xi}^T\rangle - 2\langle \hat{\Xi}^T\Theta^{(a)}\hat{\Xi}\hat{U}^{(a)}\hat{\Lambda}\hat{D}^{(a)}, U^{(a)} - \hat{U}^{(a)}\rangle + \frac{1}{2}\|\hat{\Lambda}\hat{D}^{(a)}\|_F^2 \right\} \\
&+ \sum_{a=1}^M \sum_{k=1}^{K_X} \left\{ \frac{\lambda_k}{2}\|U_k^{(a)} - e_k\|_2^2 + \frac{\nu_k}{2}(\hat{D}_{kk}^{(a)} - \hat{b}^{(a)})^2 \right\} + \frac{\omega}{2}\mathrm{tr}(\hat{\Xi}^T B_0 \hat{\Xi}).
\end{aligned}
$$

Algorithm 1 summarizes the procedure for solving the problem (C.3).

Algorithm 1 Common feature estimation under the model (C.3)

Input: $\Theta^{X(a)}$, $a = 1, \ldots, M$, and hyperparameter ρ^τ
Output: $\Xi^K,\ \Lambda^K,\ \{D^{(a)K}, U^{(a)K}\}_{a=1}^M, b^K$
1. Initialize parameters $\Xi^0,\ \Lambda^0,\ \{D^{(a)0}, U^{(a)0}\}_{a=1}^M, b^0$ using Algorithm 2
2. For $\tau = 0$ to $\tau = K - 1$
   $\Xi^{\tau+1} = \arg\min_{\Xi^T\Xi = I} \hat{L}_1(\Xi; \Xi^\tau, \Lambda^\tau, \{D^{(a)\tau}, U^{(a)\tau}\}_{a=1}^M, b^\tau) + \frac{\rho^\tau}{2}\|\Xi - \Xi^\tau\|_F^2$
   $\Lambda^{\tau+1} = \arg\min_{\Lambda \succeq 0} L(\Xi^{\tau+1}, \Lambda, \{D^{(a)\tau}, U^{(a)\tau}\}_{j=1}^M, b^\tau) + \frac{\rho^\tau}{2}\|\Lambda - \Lambda^\tau\|_F^2$
   $\{D^{(a)\tau+1}\}_{a=1}^M = \arg\min_{D^{(a)} \succeq 0} L(\Xi^{\tau+1}, \Lambda^{\tau+1}, \{D^{(a)}, U^{(a)\tau}\}_{a=1}^M, b^\tau) + \frac{\rho^\tau}{2}\sum_{a=1}^M \|D^{(a)} - D^{(a)\tau}\|_F^2$
   $b^{\tau+1} = \arg\min_{1^T b = M,\ b \geq 0} L(\Xi^{\tau+1}, \Lambda^{\tau+1}, \{D^{(a)\tau+1}, U^{(a)\tau}\}_{a=1}^M, b) + \frac{\rho^\tau}{2}\sum_{a=1}^M (b^{(a)} - b^{(a)\tau})^2$
   $\{U^{(a)\tau+1}\}_{a=1}^M = \arg\min_{(U^{(a)})^T U^{(a)} = I} \hat{L}_2(\{U^{(a)}\}_{a=1}^M; \Xi^{\tau+1}, \Lambda^{\tau+1}, \{D^{(a)\tau+1}, U^{(a)\tau}\}_{a=1}^M, b^{\tau+1})$
   $\qquad + \frac{\rho^\tau}{2}\sum_{j=1}^M \|U^{(a)} - U^{(a)\tau}\|_F^2$

In Ξ-update of Algorithm 1, the subproblem to be solved is

$$
\Xi^{\tau+1} = \arg\min_{\Xi^T\Xi = I} -2\langle A, \Xi\rangle = \arg\min_{\Xi^T\Xi = I} \|\Xi - A\|_F^2,
$$

where $A = \sum_{a=1}^M \Theta^{(a)}\Xi^\tau U^{(a)\tau}\Lambda^\tau D^{(a)\tau}(U^{(a)\tau})^T - \omega B_0 \Xi^\tau/2 + \Xi^\tau\rho^\tau/2$. Let $A = \tilde{Q}\tilde{D}\tilde{P}^T$ be the SVD decomposition of A. Then the solution is $\Xi^{\tau+1} = \tilde{Q}\tilde{P}^T$.

The subproblem for Λ-update of Algorithm 1 is

$$
\begin{aligned}
\Lambda^{\tau+1} =& \arg\min_{\Lambda \succeq 0} \sum_{a=1}^M \left\{ -\langle \Theta^{(a)}, \Xi^{\tau+1}U^{(a)\tau}\Lambda D^{(a)\tau}(U^{(a)\tau})^T(\Xi^{\tau+1})^T\rangle + \frac{1}{2}\|\Lambda D^{(a)\tau}\|_F^2 \right\} \\
&+ \frac{\rho^\tau}{2}\|\Lambda - \Lambda^\tau\|_F^2 \\
=& \arg\min_{\Lambda \succeq 0} -\mathrm{tr}(\Lambda R) + \frac{1}{2} tr\left(\Lambda^2 E\right) \\
=& \max(\mathrm{diag}(R_{11}/E_{11}, \ldots, R_{K_X K_X}/E_{K_X K_X}), 0),
\end{aligned}
$$

where $R = \sum_{a=1}^{M} D^{(a)\tau}(U^{(a)\tau})^T(\Xi^{\tau+1})^T\Theta^{(a)}\Xi^{\tau+1}U^{(a)\tau} + \rho^\tau\Lambda^\tau$ and $E = \sum_{a=1}^{M}(D^{(a)\tau})^2 + \rho^\tau I$.

It can be proved that D-update of Algorithm 1 consists of M decoupled subproblems:

$$\begin{aligned}
D^{(a)\tau+1} &= \arg\min_{D^{(a)}\succeq 0} -\langle\Theta^{(a)}, \Xi^{\tau+1}U^{(a)\tau}\Lambda^{\tau+1}D^{(a)}(U^{(a)\tau})^T(\Xi^{\tau+1})^T\rangle + \frac{1}{2}\|\Lambda^{\tau+1}D^{(a)}\|_F^2 \\
&\quad + \sum_{k=1}^{K_X}\frac{\nu_k}{2}(D_{kk}^{(a)} - b^{(a)\tau})^2 + \frac{\rho^\tau}{2}\|D^{(a)} - D^{(a)\tau}\|_F^2 \\
&= \arg\min_{D^{(a)}\succeq 0} -\text{tr}(D^{(a)}F) + \frac{1}{2}\text{tr}((D^{(a)})^2G) \\
&= \max(\text{diag}(F_{11}/G_{11}, \ldots, F_{K_XK_X}/G_{K_XK_X}), 0), \quad a = 1, \ldots, M,
\end{aligned}$$

where $F = (U^{(a)\tau})^T(\Xi^{\tau+1})^T\Theta^{(a)}\Xi^{\tau+1}U^{(a)\tau}\Lambda^{\tau+1} + b^{(a)\tau}\text{diag}(\nu_1, \ldots, \nu_{K_X}) + \rho^\tau D^{(a)\tau}$ and $G = (\Lambda^{\tau+1})^2 + \text{diag}(\nu_1, \ldots, \nu_{K_X}) + \rho^\tau I$.

Let $\nu = \text{diag}(\nu_1, \ldots, \nu_{K_X})$, $F_b = (\sum_{k=1}^{K_X}\nu_k)I$ and $\psi = (\text{tr}(\nu D^{(1)\tau+1}), \ldots, \text{tr}(\nu D^{(M)\tau+1}))^T$. The subproblem for b-update of Algorithm 1 is

$$b^{\tau+1} = \arg\min_{1^Tb=M,\ b\geq 0} \frac{1}{2}b^T(F_b + \rho^\tau\mathbf{I})b - (\psi + \rho^\tau b^\tau)^Tb,$$

which is a quadratic programming problem.

U-update of Algorithm 1 also consists of M decoupled subproblems:

$$\begin{aligned}
U^{(a)\tau+1} &= \arg\min_{(U^{(a)})^TU^{(a)}=I} -2\langle(\Xi^{\tau+1})^T\Theta^{(a)}\Xi^{\tau+1}U^{(a)\tau}\Lambda^{\tau+1}D^{(a)\tau+1}, U^{(a)}\rangle + \sum_{k=1}^{K_X}\frac{\lambda_k}{2}\|U_k^{(a)} - e_k\|_2^2 \\
&\quad + \frac{\rho^\tau}{2}\|U^{(a)} - U^{(a)\tau}\|_F^2 \\
&= \arg\min_{(U^{(a)})^TU^{(a)}=I} -2\langle H, U^{(a)}\rangle = \arg\min_{(U^{(a)})^TU^{(a)}=I}\|H - U^{(a)}\|_F^2, \quad a = 1, \ldots, M,
\end{aligned}$$

where $H = (\Xi^{\tau+1})^T\Theta^{(a)}\Xi^{\tau+1}U^{(a)\tau}\Lambda^{\tau+1}D^{(a)\tau+1} + U^{(a)\tau}\rho^\tau/2 + \text{diag}(\lambda_1, \ldots, \lambda_{K_X})/2$. Let $H = \tilde{Q}\tilde{D}\tilde{P}^T$ be the SVD decomposition of H. Then the solution is $U^{(a)\tau+1} = \tilde{Q}\tilde{P}^T$.

In practice, we set $\omega = 0.01/\text{tr}\left((\Xi^0)^TB_0\Xi^0\right)$, and choose proportional tuning parameters λ_k and ν_k by $\lambda_k = a_0\Lambda_{kk}^0$ and $\nu_k = a_0\Lambda_{kk}^0$, where Λ_{kk}^0 denotes the (k,k)-element of the initial Λ^0. The scaling constant a_0 is selected via cross-validation (Appendix D).

D Cross-validation for tuning parameter a_0

We provide detailed cross-validation steps for tuning the common penalty a_0 for X part estimation in Algorithm 3. For each source a, create K folds datasets

$$T_{a,r}\ (\text{train}), \qquad V_{a,r}\ (\text{valid}), \qquad r = 1, \ldots, K,$$

with $|T_{a,r}|, |V_{a,r}| \geq 1$.

To reduce computational cost, we may select a_0 using only a single validation split. For each source a, we randomly partition the sample into a 90% training set and a 10% validation set. The validation error for each candidate $a_0 \in A_0$ is then computed following the steps in Algorithm 3.

Algorithm 2 Parameter initialization for the model (C.3)

Input: $\Theta^{X(a)}$, $a = 1, \ldots, M$
Output: Intialized parameters $\Xi^0, \Lambda^0, \{D^{(a)0}, U^{(a)0}\}_{a=1}^M$, b^0
1. $\bar{\Theta} = \sum_{a=1}^M \Theta^{X(a)}/M$.
2. Let $\bar{\Theta} = \tilde{\Xi}\tilde{\Lambda}(\tilde{\Xi})^T$ be the eigendecomposition of $\bar{\Theta}$. Initialize $\Xi^0 = \tilde{\Xi}$.
3. For $a = 1$ to $a = M$
   Let $\Theta^{X(a)} = \tilde{\Xi}^{(a)}\tilde{\Lambda}^{(a)}(\tilde{\Xi}^{(a)})^T$ be the eigendecomposition of $\Theta^{X(a)}$.
   Let $\tilde{\Xi}^T\tilde{\Xi}^{(a)} = \tilde{Q}^{(a)}\tilde{D}^{(a)}(\tilde{P}^{(a)})^T$ be the SVD decomposition of $\tilde{\Xi}^T\tilde{\Xi}^{(a)}$.
   Let $\text{diag}(\tilde{D}^{(a)0}) = \text{diag}(\tilde{\Lambda}^{(a)})/\text{diag}(\tilde{\Lambda})$.
   Initialize $U^{(a)0} = \tilde{Q}^{(a)}(\tilde{P}^{(a)})^T$.
4. Let $c = \sum_{a=1}^M \tilde{D}_{11}^{(a)0}/M$.
5. Initialize $D^{(a)0} = \tilde{D}^{(a)0}/c$, $a = 1, \ldots, M$, $b^0 = (D_{11}^{(1)0}, \ldots, D_{11}^{(M)0})^T$.
6. Initialize $\Lambda^0 = c\tilde{\Lambda}$.

Algorithm 3 K-fold cross-validation for selecting a_0

Input: candidate $a_0 \in A_0$; within–source folds $\{T_{a,r}, V_{a,r}\}_{a=1,\ldots,M}^{r=1,\ldots,K}$
Output: optimal penalty $a_0^\star$
For $r = 1, \ldots, K$
   Estimate $\alpha(t), \gamma^{(a)}(t), \beta(t)$ in (1) using $T_{a,r}$ and get estimates $\hat{\alpha}_r(t), \hat{\gamma}_r^{(a)}(t), \hat{\beta}_r(t)$.
   $\hat{\eta}_{ij}^{(a)}(t) = \tilde{Y}_{ij}^{(a)}(t) - \hat{\alpha}_r(t) - \hat{\gamma}_r^{(a)}(t) - (Z_{ij}^{(a)})^T\hat{\beta}_r(t)$.
   $\Theta_{\text{train},r}^{X(a)} = \text{MFPCA}\big(\{\hat{\eta}_{ij}^{(a)}\}_{i\in T_{a,r}}\big), \quad \Theta_{\text{val},r}^{X(a)} = \text{MFPCA-reconstruct}\big(\{\hat{\eta}_{ij}^{(a)}\}_{i\in V_{a,r}}\big)$.
   For $a_0 \in A_0$
      Plug $\Theta_{\text{train},r}^{X(a)}$ in (C.3) and obtain $\hat{\Xi}_r^X$ and $\hat{\Lambda}_r^X$.
      Estimate $\{U_{\text{val}}^{(a)},\ D_{\text{val}}^{(a)}, b_{\text{val}}^{(a)}\}$ by solving
         minimize $\big\|\Theta_{\text{val},r}^{X(a)} - \hat{\Xi}_r^X U\hat{\Lambda}_r^X DU^\top\hat{\Xi}_r^{X\top}\big\|_F^2 + a_0\Lambda_{kk}^{X0}\sum_{k=1}^{K_X}(\|U_k - e_k\|_2^2$
         $+(D_{kk} - b^{(a)})^2)$
         subject to $1^T b = M,\ b^{(a)} \geq 0$.
      Fold error $E_r(a_0) = \frac{1}{M}\sum_{a=1}^M\big\|\Theta_{\text{val},r}^{X(a)} - \hat{\Xi}_r^X\hat{U}_{\text{val}}^{(a)}\hat{\Lambda}_r^X\hat{D}_{\text{val}}^{(a)}\hat{U}_{\text{val}}^{(a)\top}\hat{\Xi}_r^{X\top}\big\|_F^2$.
Average over folds $\text{CV}(a_0) = \dfrac{1}{K}\displaystyle\sum_{r=1}^K E_r(a_0)$.
Optimal penalty $a_0^\star = \arg\min_{a_0\in A_0} \text{CV}(a_0)$.

E Simple standardization and quantile normalization approaches

To ensure comparability with INTACT and INTACT$_0$, both simple standardization and quantile normalization are applied to the raw data at each timepoint.

E.1 Simple standardization

For each t, let $\mathcal{A}^{(a)}(t) = \{\hat{\eta}_{ij}^{(a)} : i = 1, \ldots, n_a,\ j = 1, \ldots, r_i\}$ denote the set of $\hat{\eta}_{ij}^{(a)}$ in source a. Let $e(t)$ be the variance of the pooled data $\cup_a \mathcal{A}^{(a)}(t)$, and $e^{(a)}(t)$ the variance within $\mathcal{A}^{(a)}(t)$. The standardized data are then defined as

$$\hat{Y}_{ij}^{s(a)}(t) = c_0\big(\hat{\alpha}(t) + Z_{ij}^T\hat{\beta}(t) + \hat{\eta}_{ij}^{s(a)}(t)\big),$$

where $\hat{\eta}_{ij}^{s(a)}(t) = \hat{\eta}_{ij}^{(a)}(t)\sqrt{e(t)/e^{(a)}(t)}$.

E.2 Quantile normalization

For each t, we apply the `normalize` function in the R package `broman` to $(\mathcal{A}^{(1)}(t), \ldots, \mathcal{A}^{(M)}(t))$. If the lengths of $\mathcal{A}^{(1)}(t), \ldots, \mathcal{A}^{(M)}(t)$ differ, they are first equalized via linear interpolation. After normalization, the results are interpolated back to the original lengths, yielding $\hat{\eta}_{ij}^{q(a)}(t)$. The quantile-normalized data are defined as

$$\hat{Y}_{ij}^{q(a)}(t) = c_0\big(\hat{\alpha}(t) + Z_{ij}^T\hat{\beta}(t) + \hat{\eta}_{ij}^{q(a)}(t)\big).$$

F Additional simulation results

F.1 Definition of basic 3D rotation matrices

The definitions of basic 3D rotation matrices are as follows:

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix},$$

$$R_y(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix},$$

$$R_z(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

F.2 Heatmaps of covariance matrices and source-specific rotations and scales

Figure 5 presents source-specific rotations and scales.

F.3 Simulation results for incomplete data settings

Figure 6 and Table 6 present the results for the incomplete data settings. The findings closely mirror those obtained under the complete data scenario.

Table 3: Simulation results for incomplete data, including mean computation time (Time, seconds); mean variance discrepancies (MV^X, DV^X, MV^W, DV^W); empirical power for detecting source effects (Source-effect); and mean squared errors (C1–C3 MSE) and empirical power (C1–C3 Power) for β_2 across three Cox models.

	raw	standard	quantile	L-ComBat	INTACT	INTACT_0
Time (seconds)	0.672	2.238	2.315	2.194	2.854	2.854
MV^X	0.714	0.566	0.560	0.760	0.518	0.447
DV^X	2.123	1.512	1.338	2.375	0.057	0.064
MV^W	0.227	0.232	0.353	0.241	0.185	0.156
DV^W	0.547	0.561	0.457	0.528	0.003	0.003
Source-effect	0.680	0.715	0.995	0.655	0.115	0.040
C1 MSE	0.462	0.412	0.434	0.409	0.391	0.390
C2 MSE	0.006	0.005	0.005	0.005	0.005	0.004
C3 MSE	12.367	10.884	11.960	10.863	10.769	10.662
C1 Power	0.192	0.200	0.225	0.223	0.225	0.228
C2 Power	0.550	0.742	0.725	0.750	0.767	0.767
C3 Power	0.783	0.976	0.958	0.983	1.000	1.000

G Numerical results when the shared-space assumption is violated

We evaluate the proposed method when the shared-space assumption does not hold. Data are generated from

$$Y_{ij}^{(a)}(t) = \alpha(t) + \gamma^{(a)}(t) + Z_i^{(a)}\beta_1(t) + \eta_{ij}^{(a)}(t), \quad a = 1, \ldots, M, \quad i = 1, \ldots, n_a, \quad j = 1, \ldots, r_i,$$

where

$$\eta_{ij}^{(a)}(t) = \Psi^{X(a)}(t)(\Lambda^X)^{1/2}(D^{X(a)})^{1/2}\xi_i^{X(a)} + \Psi^{W(a)}(t)(\Lambda^W)^{1/2}(D^{W(a)})^{1/2}\xi_{ij}^{W(a)} + \epsilon_{ij}^{(a)}(t),$$

with $t \in \{1, \ldots, 144\}$.

We set $M = 2$, $\alpha(t) = \gamma^{(a)}(t) = 0$, $n_a = 200$, and $r_i \in \{3, 4, 5, 6, 7\}$. Let

$$\Lambda^X = \text{diag}(10, 5, 1),\ D^{X(1)} = \text{diag}(0.9, 0.8, 0.8),\ D^{X(2)} = \text{diag}(1.1, 1.2, 1.2),$$
$$\Lambda^W = \text{diag}(0.5, 0.3, 0.1),\ D^{W(1)} = \text{diag}(1.2, 1.1, 1.1),\ D^{W(2)} = \text{diag}(0.8, 0.85, 0.85).$$

We generate $Z_i^{(a)} \overset{\text{i.i.d.}}{\sim} N(1, 0.5^2)$, $\xi_i^{X(a)} \overset{\text{i.i.d.}}{\sim} N(0, I_3)$, $\xi_{ij}^{W(a)} \overset{\text{i.i.d.}}{\sim} N(0, I_3)$, and $\epsilon_{ij}^{(a)}(t) \overset{\text{i.i.d.}}{\sim} N(0, 0.2^2)$, for $a = 1, 2$. The shape of $\beta_1(t)$ is depicted by the red curve in Figure 7B (and Figure 8B).

Two scenarios are considered: (i) source-specific eigenfunctions are reasonably aligned but not in a common space; (ii) source-specific eigenfunctions differ substantially. Shapes of $\Psi^{X(a)}(t)$ and $\Psi^{W(a)}(t)$ are shown in Figures 7A and 8A, respectively.

For comparability, we scale each $Y_{ij}^{(a)}(t)$ by the square root of the spectral norm of its sample covariance matrix (Step 1, Section 2.2). The scaled data $\tilde{Y}_{ij}^{(a)}(t)$ serve as the raw baseline. We compare INTACT with simple standardization, quantile normalization (Amaratunga and Cabrera, 2001; Bolstad et al., 2003), and longitudinal ComBat (Beer et al., 2020), generating 200 datasets under each scenario. Following Section 3, we assess harmonization effectiveness and test $H_0 : \beta_1(t) = 0$ at each t.

For Scenario 1 (Figure 7 and the first row of Table 4), the estimated $\hat{\beta}_1(t)$ closely matches the truth (Figure 7B). INTACT and INTACT_0 achieve higher power in detecting biological effects (Figure 7C),

Table 4: Empirical power for detecting source effects when integrating data with nonshared eigenfunction spaces.

	raw	standard	quantile	L-ComBat	INTACT	INTACT_0
Scenario 1	1	1	1	1	0.145	0.095
Scenario 2	1	1	1	1	0.130	0.100

and yield lower AUC/ARI values (Figure 7D), indicating reduced source separability. Table 4 shows INTACT and INTACT_0 have substantially lower empirical power for detecting source effects.

For Scenario 2 (Figure 8 and the second row of Table 4), results for $\beta_1(t)$ estimation and source-effect detection are similar to Scenario 1. However, the empirical power for detecting $\beta_1(t)$ (Figure 8C) drops notably around $t \approx 100$, and in some regions, raw1/raw2 outperform all harmonization methods—both undesirable outcomes.

Overall, when the common eigenspace assumption is violated, numerical results show that INTACT performs well when source-specific eigenspaces are reasonably aligned. When eigenfunctions differ substantially, INTACT still effectively removes source effects, but the preservation of biological signals may be compromised.

H Additional results for integration of NHANES 2003-2004 and 2013-2014 data

H.1 Heatmaps of covariance matrices and source-specific rotations and scales

The heatmaps of covariance for raw and harmonized data are shown in Figure 9. Figure 10 presents source-specific rotations and scales.

H.2 Choosing the number of components with different weights

Figure 11A presents the curves of $\text{mPVE}^X(k)$, $\text{Sim}^X(k)$, $\text{mPVE}^W(k)$, and $\text{Sim}^W(k)$. Based on these curves, we select the numbers of eigenfunctions K_X and K_W under various choices of α^X and α^W (Table 5(a)), yielding five scenarios: (S1) $K_X = K_W = 2$, (S2) $K_X = 3,\ K_W = 2$, (S3) $K_X = 3,\ K_W = 26$, (S4) $K_X = 5,\ K_W = 26$, and (S5) $K_X = 9,\ K_W = 26$.

We apply the proposed methods under each scenario and assess both source effects and age effects. Figure 11B displays heatmaps of covariance differences for the X and W parts. Overall, the proposed methods yield smaller differences than alternative harmonization approaches, with INTACT and INTACT_0 in scenario S2 performing best.

To quantify source effects, we repeatedly (200 times) resample 150 subjects from each source, select one observation (day) per subject, and remove the estimated fixed effects $\hat{\alpha}(t) + Z_{ij}^{(a)T}\hat{\beta}(t)$ from the harmonized data. For each resampled dataset, we apply the generalized graph-based test and the random forest classifier, as in the simulation studies. Table 5(b) reports the empirical power for testing equality of distributions at the 0.05 significance level. The proposed methods generally achieve lower power, indicating better harmonization; scenario S2 yields the smallest power for both INTACT and INTACT_0. Figure 11C presents boxplots of AUC and ARI, showing that INTACT and INTACT_0—especially in S3–S5—produce the lowest values, with quantile normalization reducing but not eliminating source effects.

Finally, to assess whether harmonization strengthens the association between age and physical activity patterns, we test the null hypothesis of no age effect on activity intensity. In each of 200 repetitions,

Table 5: Integration results for NHANES 2003–2004 and NHANES 2013–2014 accelerometer data.

(a) Number of eigenfunctions K_X and K_W under different choices of α^X and α^W.

α^X (α^W)	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
K_X	2	2	2	3	3	3	3	3	5	9	9
K_W	2	2	2	2	2	2	26	26	26	26	26

(b) Empirical power for detecting source effects across different harmonization methods and scenarios.

raw	standard	quantile	L-ComBat	(S1)INTACT	(S1)INTACT$_0$
0.915	0.690	1	0.785	0.650	0.095
(S2)INTACT	(S2)INTACT$_0$	(S3)INTACT	(S3)INTACT$_0$	(S4)INTACT	(S4)INTACT$_0$
0.550	0.090	0.620	0.630	0.675	0.625
(S5)INTACT	(S5)INTACT$_0$				
0.680	0.645				

we resample 150 subjects per source and apply a pointwise empirical likelihood ratio test (Wang et al., 2010) to a linear mixed-effects model. Figure 11D shows that INTACT and INTACT$_0$ consistently exhibit higher power than existing methods across nearly the entire time range.

In conclusion, the proposed method outperforms alternative harmonization approaches across different weight choices in criterion (6), though excessively small or large weights may amplify source effects. In the absence of prior knowledge regarding the number of components, we recommend using criterion (6) with weights between 0.3 and 0.7.

H.3 Association between mortality and minute-level physical activity

We assess the association between mortality and minute-level physical activity by fitting a Cox proportional hazards model to log-transformed activity intensities, adjusted for gender, age, income, and BMI, using a 5-year censoring period. The model is implemented via the `coxph()` function in the R package `survival`. For participants with multiple days of data, we average the log-transformed intensities across days. Separate Cox models are then fitted to the activity intensity at each minute.

Figure 12 presents the hazard ratios and Bonferroni-corrected 99.997% (=1-0.05/1440) confidence intervals for each minute. Across all methods, hazard ratios fall below 1 from approximately 8am to 10pm, indicating that higher activity during these hours is associated with reduced mortality risk. Between 12am and 6am, INTACT and INTACT$_0$ yield hazard ratios above 1, whereas other methods produce estimates near or slightly above 1.

I Integration of NHANES accelerometer and ExtraSensory gyroscopes data

I.1 Parameter estimation and source effect evaluation

The first row of Table 6 summarizes computation times, showing that longitudinal ComBat is the most time-consuming method, whereas all others complete within 15 seconds. Figure 13B presents the estimated common mean curve and source-specific mean shift functions, revealing lower nighttime and higher daytime activity. Using the selection criterion in (6), we retain 8 eigenfunctions for the X part

Table 6: Computation time (Time, seconds) for integrating NHANES accelerometer and ExtraSensory gyroscope data, and empirical power for detecting TAC and TLAC differences between two sleep onset groups.

	raw	standard	quantile	L-ComBat	INTACT	INTACT_0
Time (seconds)	0.647	7.507	11.472	43.751	14.382	14.382
TAC	0.524	0.140	0.992	0.136	1.000	1.000
TLAC	0.088	0.098	0.258	0.208	0.996	1.000

and 12 for the W part. Figure 13C displays the first three eigenfunctions. Figure 13C shows the first three eigenfunctions for both X and W parts. For the X part, the first eigenfunction reflects elevated daytime activity; the second is positive throughout the day except from 6 am to 9 am, indicating reduced early-morning activity; and the third captures reduced activity during working hours (9 am–6 pm) with elevated activity otherwise. The W part represents within-subject variation: the first eigenfunction corresponds to higher activity from midnight to noon; the second indicates reduced activity from midnight to 8 am; and the third reflects lower morning activity relative to the individual's average pattern.

We next assess source effects in the harmonized data. Figure 13D displays the overall mean curves, showing that harmonization markedly reduces mean differences between sources compared to the raw data. We estimate sample covariances using fast MFPCA, and Figure 13E presents heatmaps of the covariance differences for the X and W parts. Quantile normalization, INTACT, and INTACT_0 produce minimal covariance discrepancies, whereas the raw data and other methods exhibit larger differences in covariances.

I.2 Association between sleep onset time and physical activity

Previous studies indicate that sleep onset time influences physical activity patterns (Master et al., 2019; Leota et al., 2025). To assess whether harmonization improves detection of such associations, we test the null hypothesis of no sleep onset effect on daily activity intensity. Minute-level sleep annotations classify each day as before or after midnight sleep onset. One day per subject is randomly selected, yielding 786 days in Group 1 (before 12 am) and 489 days in Group 2 (after 12 am). Daily activity intensities are summarized by TAC and TLAC, and compared between groups using the Wilcoxon test (`wilcox.test()` in `stats`). This procedure is repeated 200 times, and the second and third rows of Table 6 report empirical power at the 0.05 level. INTACT and INTACT_0 achieve markedly higher power than competing methods, demonstrating greater sensitivity for detecting sleep onset associations.

References

Amaratunga, D. and J. Cabrera (2001). Analysis of data from viral dna microchips. *Journal of the American Statistical Association 96*(456), 1161–1170.

Beer, J. C., N. J. Tustison, P. A. Cook, C. Davatzikos, Y. I. Sheline, R. T. Shinohara, K. A. Linn, A. D. N. Initiative, et al. (2020). Longitudinal combat: A method for harmonizing longitudinal multi-scanner imaging data. *Neuroimage 220*, 117129.

Benito, M., J. Parker, Q. Du, J. Wu, D. Xiang, C. M. Perou, and J. S. Marron (2004). Adjustment of systematic microarray data biases. *Bioinformatics 20*(1), 105–114.

Bolstad, B. M., R. A. Irizarry, M. Åstrand, and T. P. Speed (2003). A comparison of normalization

methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics 19*(2), 185–193.

Breiman, L. (2001). Random forests. *Machine learning 45*, 5–32.

Butler, A., P. Hoffman, P. Smibert, E. Papalexi, and R. Satija (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology 36*(5), 411–420.

Chen, A. A., J. C. Beer, N. J. Tustison, P. A. Cook, R. T. Shinohara, H. Shou, and A. D. N. Initiative (2022). Mitigating site effects in covariance for machine learning in neuroimaging data. *Human brain mapping 43*(4), 1179–1195.

Chen, H. and J. H. Friedman (2017). A new graph-based two-sample test for multivariate and object data. *Journal of the American statistical association 112*(517), 397–409.

Crainiceanu, C. M., J. Goldsmith, A. Leroux, and E. Cui (2024). *Functional Data Analysis with R* (1st ed.). Chapman and Hall/CRC.

Cui, E., C. M. Crainiceanu, and A. Leroux (2021). Additive functional cox model. *Journal of Computational and Graphical Statistics 30*(3), 780–793.

Cui, E., R. Li, C. M. Crainiceanu, and L. Xiao (2023). Fast multilevel functional principal component analysis. *Journal of Computational and Graphical Statistics 32*(2), 366–377.

Di, C.-Z., C. M. Crainiceanu, B. S. Caffo, and N. M. Punjabi (2009). Multilevel functional principal component analysis. *The annals of applied statistics 3*(1), 458.

Hansen, B. B. (2007). Optmatch: Flexible, optimal matching for observational studies. *New Functions for Multivariate Analysis 7*(2), 18–24.

Harrison, D., P. Marshall, N. Bianchi-Berthouze, and J. Bird (2015). Activity tracking: barriers, workarounds and customisation. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 617–621.

Hastie, T., R. Mazumder, J. D. Lee, and R. Zadeh (2015). Matrix completion and low-rank svd via fast alternating least squares. *The Journal of Machine Learning Research 16*(1), 3367–3402.

Hie, B., B. Bryson, and B. Berger (2019). Efficient integration of heterogeneous single-cell transcriptomes using scanorama. *Nature biotechnology 37*(6), 685–691.

Johnson, W. E., C. Li, and A. Rabinovic (2007). Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics 8*(1), 118–127.

Leota, J., M. É. Czeisler, F. Le, D. M. Presby, E. R. Capodilupo, H. Scott, J. F. Wiley, S. P. Drummond, S. M. Rajaratnam, and E. R. Facer-Childs (2025). Sleep duration and timing are associated with next-day physical activity: Insights from two large-scale wearable sensor studies. *Proceedings of the National Academy of Sciences 122*(27), e2420846122.

Master, L., R. T. Nye, S. Lee, N. G. Nahmod, S. Mariani, L. Hale, and O. M. Buxton (2019). Bidirectional, daily temporal associations between sleep and physical activity in adolescents. *Scientific reports 9*(1), 7732.

Peng, M., Y. Li, B. Wamsley, Y. Wei, and K. Roeder (2021). Integration and transfer learning of single-cell transcriptomes via cfit. *Proceedings of the National Academy of Sciences 118*(10), e2024383118.

Sergazinov, R., A. Leroux, E. Cui, C. Crainiceanu, R. N. Aurora, N. M. Punjabi, and I. Gaynanova (2023). A case study of glucose levels during sleep using multilevel fast function on scalar regression inference. *Biometrics 79*(4), 3873–3882.

Shou, H., V. Zipunnikov, C. M. Crainiceanu, and S. Greven (2015). Structured functional principal component analysis. *Biometrics 71*(1), 247–257.

Vaizman, Y., K. Ellis, and G. Lanckriet (2017). Recognizing detailed human context in the wild from smartphones and smartwatches. *IEEE pervasive computing 16*(4), 62–74.

Wang, S., L. Qian, and R. J. Carroll (2010). Generalized empirical likelihood methods for analyzing longitudinal data. *Biometrika 97*(1), 79–93.

Welch, J. D., V. Kozareva, A. Ferreira, C. Vanderburg, C. Martin, and E. Z. Macosko (2019). Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell 177*(7), 1873–1887.

Wrobel, J., J. Muschelli, and A. Leroux (2021). Diurnal physical activity patterns across ages in a large uk based cohort: the uk biobank study. *Sensors 21*(4), 1545.

Xiao, L., L. Huang, J. A. Schrack, L. Ferrucci, V. Zipunnikov, and C. M. Crainiceanu (2015). Quantifying the lifetime circadian rhythm of physical activity: a covariate-dependent functional approach. *Biostatistics 16*(2), 352–367.

Xiao, L., V. Zipunnikov, D. Ruppert, and C. Crainiceanu (2016). Fast covariance estimation for high-dimensional functional data. *Statistics and computing 26*, 409–421.

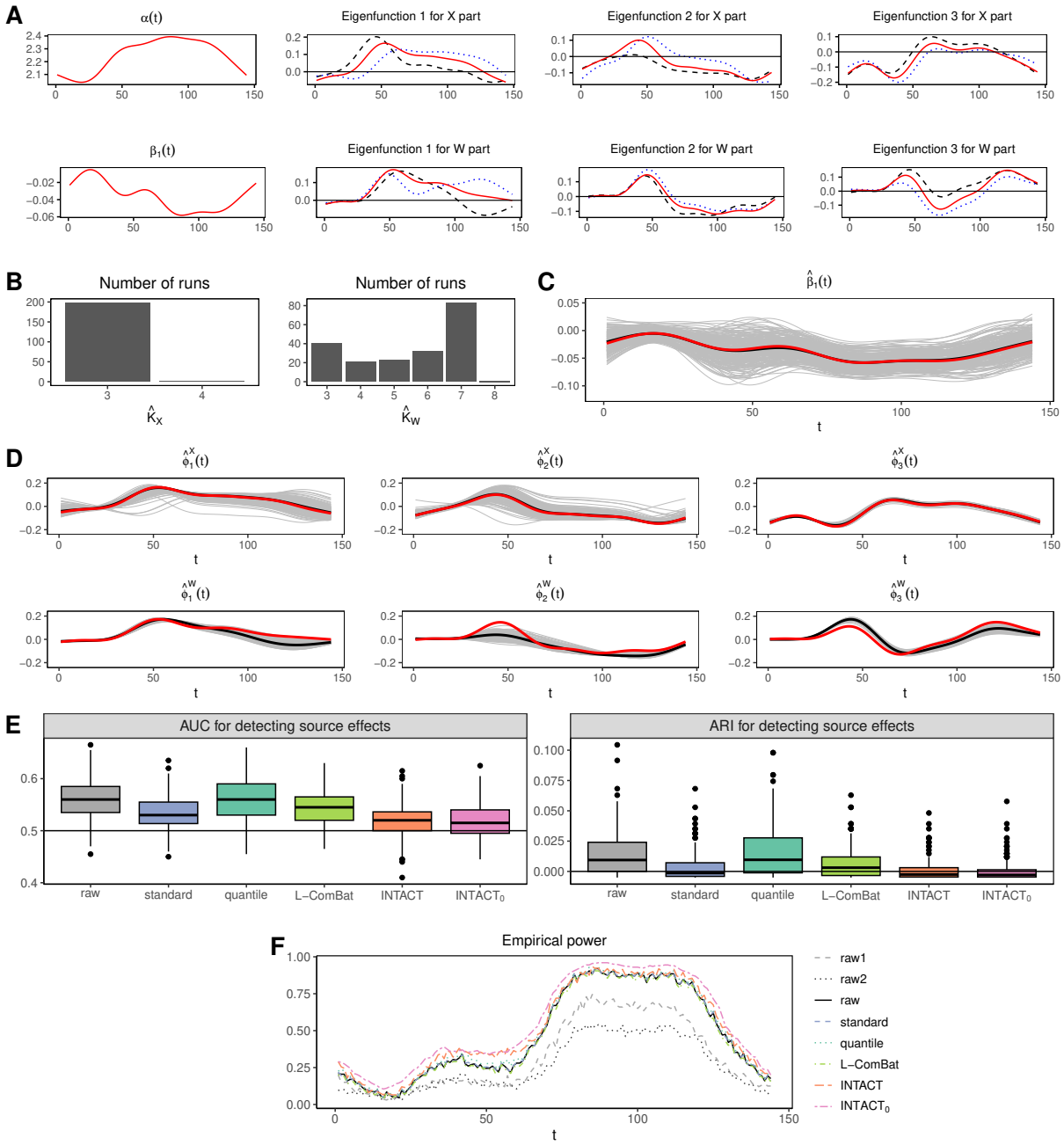

Figure 1: (A–F) Simulation settings and results for complete data. Abbreviations: source 1 data (raw1), source 2 data (raw2), combined raw data (raw), simple standardization (standard), quantile normalization (quantile), longitudinal ComBat (L-ComBat), and proposed methods (INTACT, INTACT$_0$). (A) Top: $\alpha(t)$ and eigenfunctions for the X part; bottom: $\beta_1(t)$ and eigenfunctions for the W part. In the second to fourth columns, solid red curves are the true common eigenfunctions (ϕ_k^X, ϕ_k^W); dashed black curves are $\psi_k^{X(1)}$ or $\psi_k^{W(1)}$; dotted blue curves are $\psi_k^{X(2)}$ or $\psi_k^{W(2)}$. (B) Frequency of selected numbers of eigenfunctions. (C) Estimated $\beta_1(t)$ with gray for individual runs, black for mean estimate, and red for the true function. (D) Estimated eigenfunctions with gray for individual runs, black for mean estimates, and red for the true functions. (E) Source-effect detection: left panel shows AUC values and right panel shows ARI values. (F) Empirical power for testing $H_0\colon \beta_1(t) = 0$.

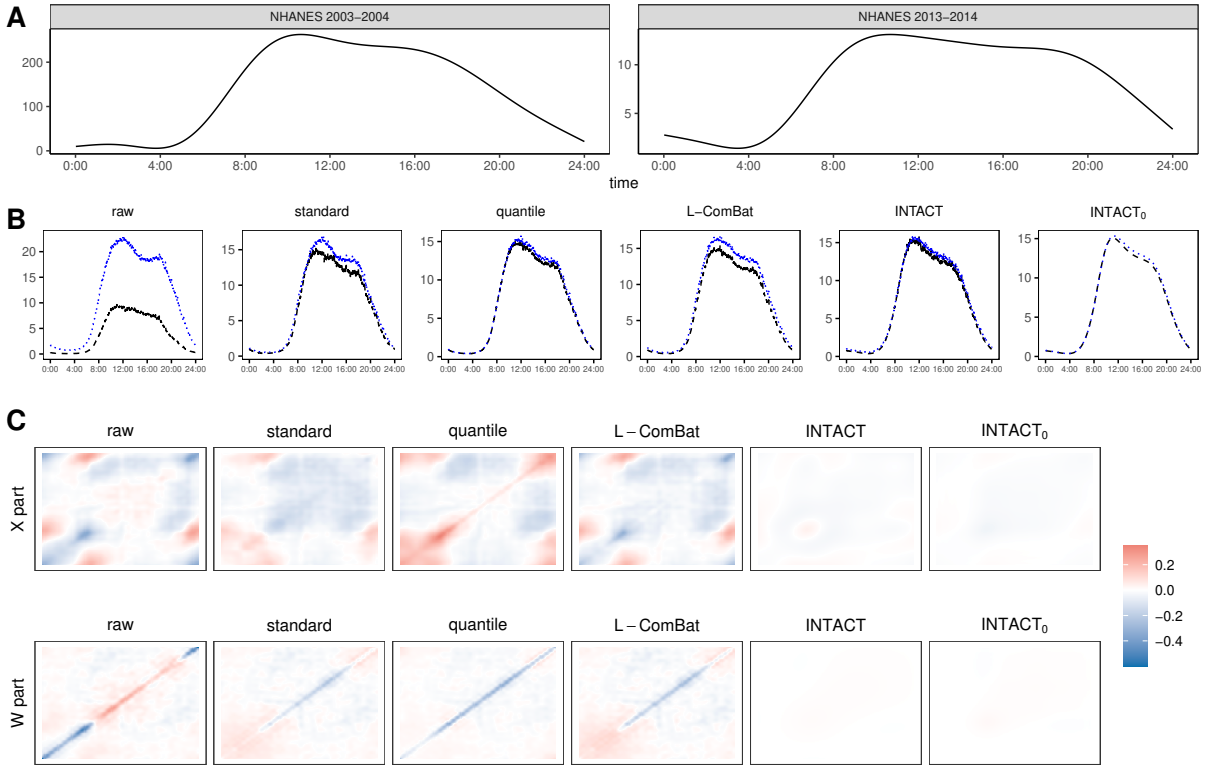

Figure 2: (A–C) Illustration of original NHANES data and mean curves and covariance differences after harmonization. (A) Mean curves of original NHANES 2003–2004 and NHANES 2013–2014 accelerometer data. (B) Mean curves of raw and harmonized data, with NHANES 2003–2004 in dashed black and NHANES 2013–2014 in dotted blue. (C) Heatmaps of covariance differences between sources.

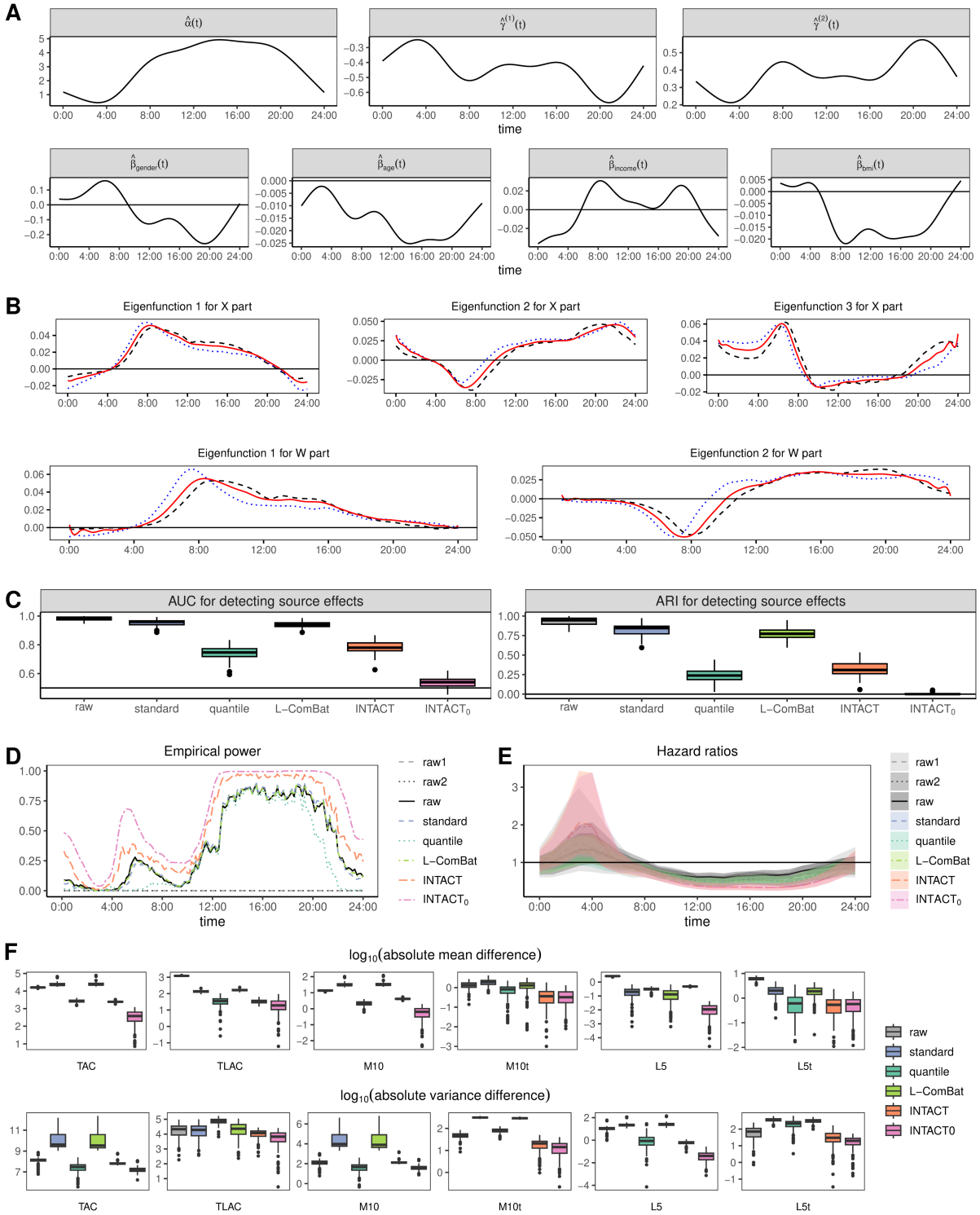

Figure 3: (A–F) Integration results for NHANES physical activity data. Abbreviations: raw data (raw), standardization (standard), quantile normalization (quantile), longitudinal ComBat (L-ComBat), proposed methods (INTACT, INTACT$_0$), NHANES 2003–2004 (raw1), NHANES 2013–2014 (raw2). (A) Estimated fixed effects. (B) First three X-part and first two W-part eigenfunctions. Dashed black: original NHANES 2003–2004; dotted blue: original NHANES 2013–2014; solid red: common eigenfunctions from INTACT. (C) Source-effect detection: AUC (left) and ARI (right). (D) Empirical power for testing no age effect. (E) Hazard ratios with Bonferroni-corrected confidence intervals from Cox models. (F) Absolute mean and variance differences in summary statistics between sources.

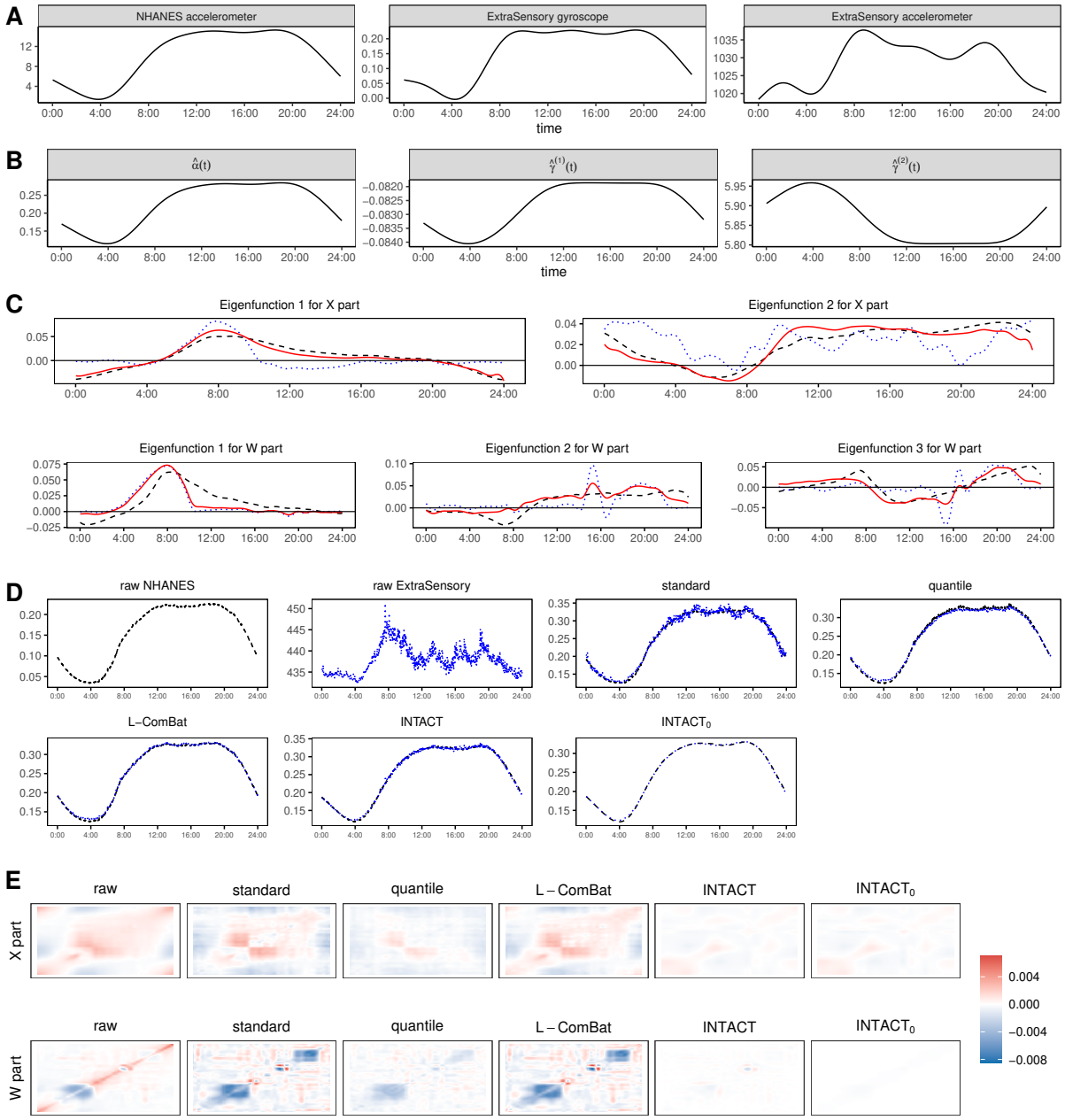

Figure 4: (A) Mean curves for original NHANES accelerometer, ExtraSensory gyroscope, and ExtraSensory accelerometer data. (B–E) Integration results for NHANES and ExtraSensory accelerometer data. Abbreviations: raw data (raw), simple standardization (standard), quantile normalization (quantile), longitudinal ComBat (L-ComBat), proposed methods (INTACT, INTACT$_0$). (B) Estimated common mean and source-specific shifts. (C) First two X-part and first three W-part eigenfunctions. Dashed black: NHANES 2013–2014; dotted blue: ExtraSensory; solid red: common eigenfunctions from INTACT. (D) Mean curves of raw and harmonized data (dashed black: NHANES 2013–2014; dotted blue: ExtraSensory). (E) Heatmaps of covariance differences between sources.

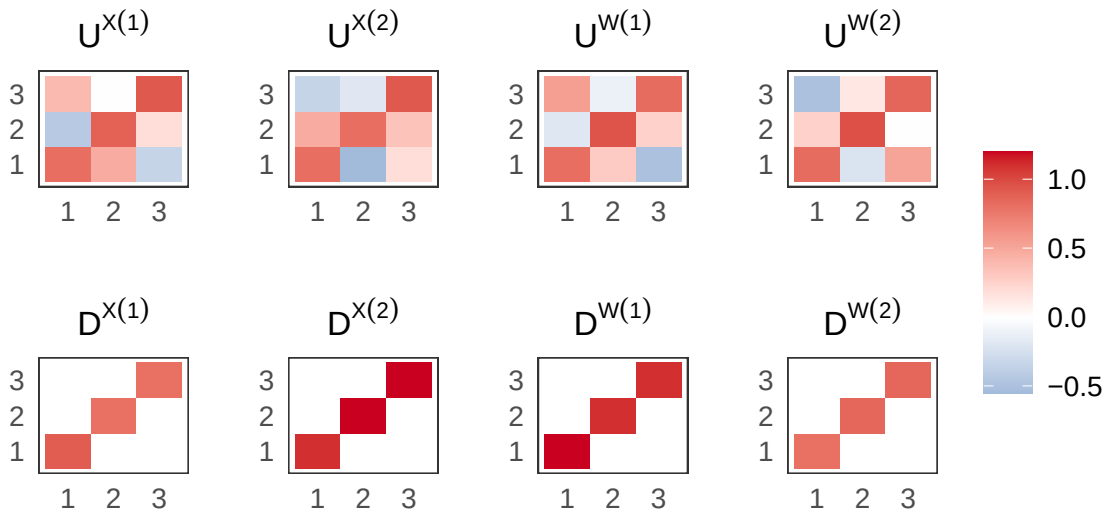

Figure 5: Heatmaps of source-specific rotations and scales in simulations.

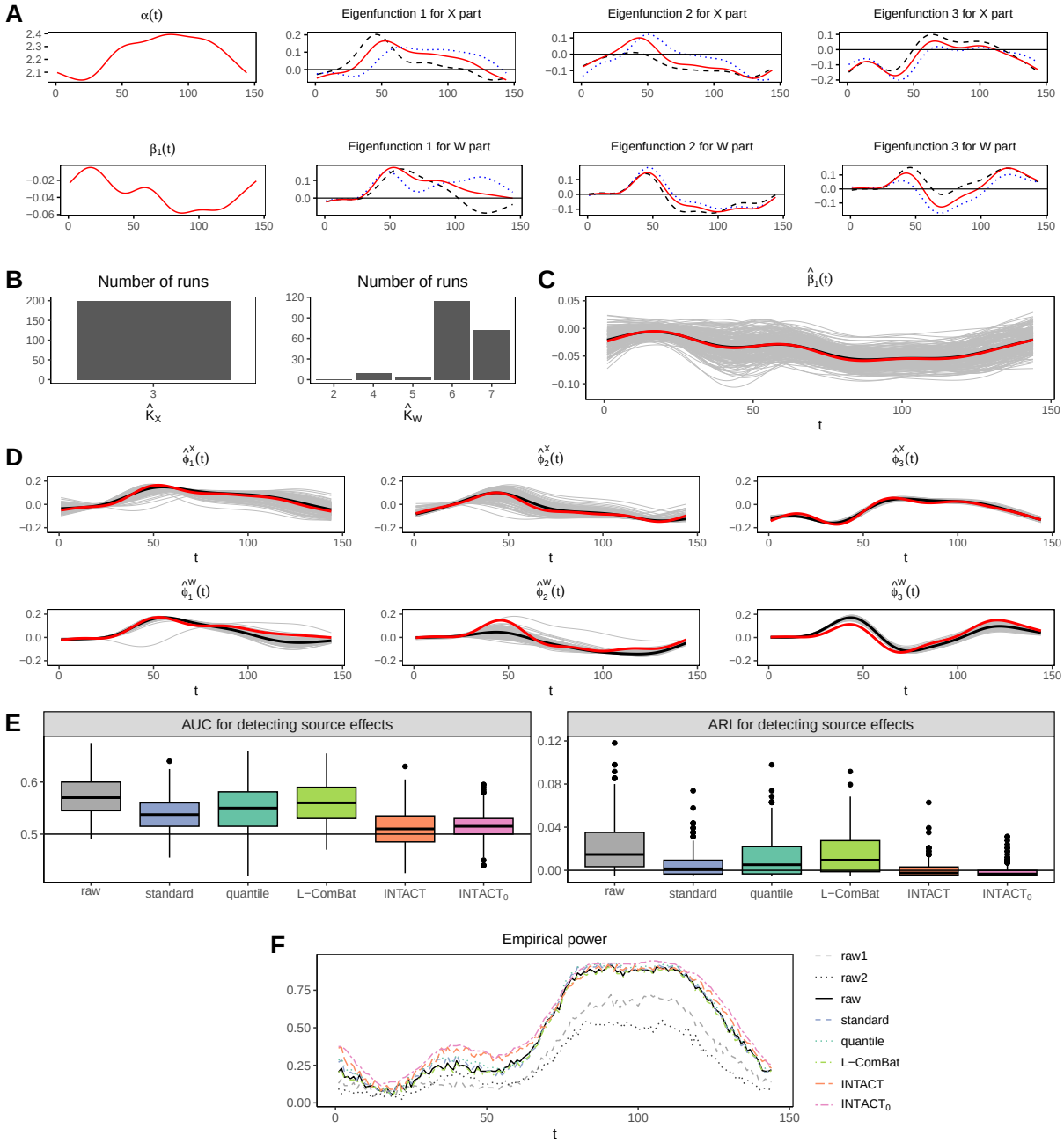

Figure 6: (A–F) Simulation settings and results for incomplete data. Abbreviations: source 1 data (raw1), source 2 data (raw2), combined raw data (raw), simple standardization (standard), quantile normalization (quantile), longitudinal ComBat (L-ComBat), and proposed methods (INTACT, INTACT$_0$). (A) Common feature functions: top showing $\alpha(t)$, $\phi_1^X(t)$, $\phi_2^X(t)$, $\phi_3^X(t)$; bottom showing $\beta_1(t)$, $\phi_1^W(t)$, $\phi_2^W(t)$, $\phi_3^W(t)$. (B) Frequency of selected numbers of eigenfunctions. (C) Estimated $\beta_1(t)$ with gray for individual runs, black for mean estimate, and red for the true function. (D) Estimated eigenfunctions with gray for individual runs, black for mean estimates, and red for the true functions. (E) Source-effect detection: left panel shows AUC values and right panel shows ARI values. (F) Empirical power for testing $H_0\colon \beta_1(t) = 0$.

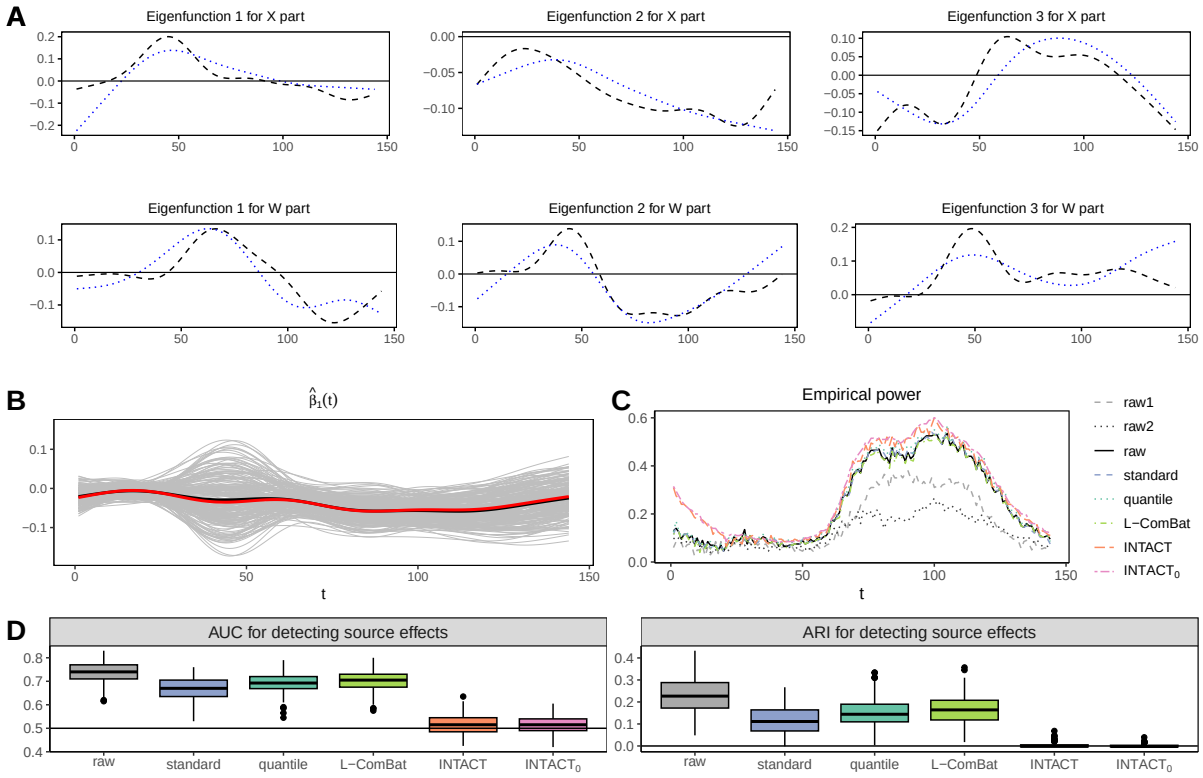

Figure 7: (A–D) Numerical results for data with similar source-specific eigenfunctions in a nonshared space. Abbreviations: source 1 (raw1), source 2 (raw2), combined raw data (raw), simple standardization (standard), quantile normalization (quantile), longitudinal ComBat (L-ComBat), proposed methods (INTACT, INTACT$_0$). (A) Source-specific eigenfunctions for the X part (top) and W part (bottom); dashed black curves: $\psi_k^{X(1)}$ or $\psi_k^{W(1)}$, dotted blue curves: $\psi_k^{X(2)}$ or $\psi_k^{W(2)}$. (B) Estimated $\beta_1(t)$ (gray: individual runs, black: mean estimate, red: true function). (C) Empirical power for testing $H_0 \colon \beta_1(t) = 0$. (D) Source-effect detection: AUC (left) and ARI (right).

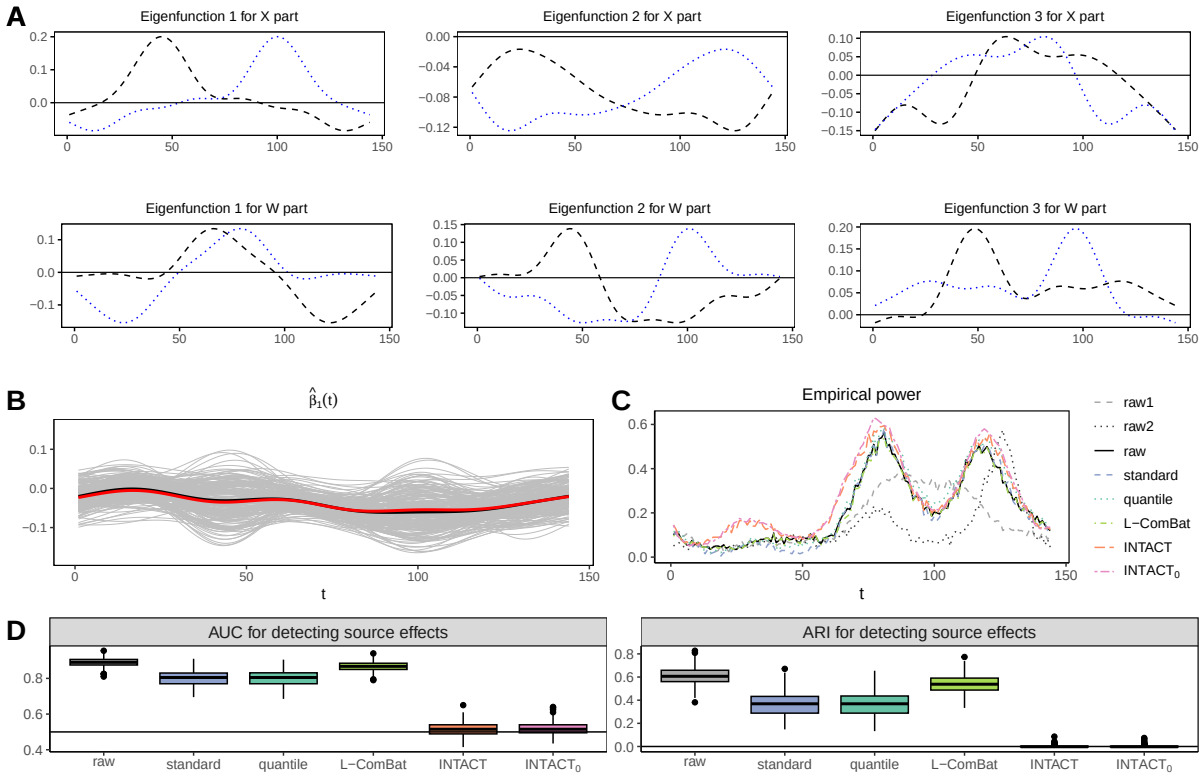

Figure 8: (A–D) Numerical results for data with markedly different source-specific eigenfunctions in a nonshared space. Abbreviations: source 1 (raw1), source 2 (raw2), combined raw data (raw), simple standardization (standard), quantile normalization (quantile), longitudinal ComBat (L-ComBat), proposed methods (INTACT, INTACT$_0$). (A) Source-specific eigenfunctions for the X part (top) and W part (bottom); dashed black: $\psi_k^{X(1)}$ or $\psi_k^{W(1)}$, dotted blue: $\psi_k^{X(2)}$ or $\psi_k^{W(2)}$. (B) Estimated $\beta_1(t)$ (gray: individual runs, black: mean, red: true function). (C) Empirical power for testing $H_0 : \beta_1(t) = 0$. (D) Source-effect detection: AUC (left) and ARI (right).

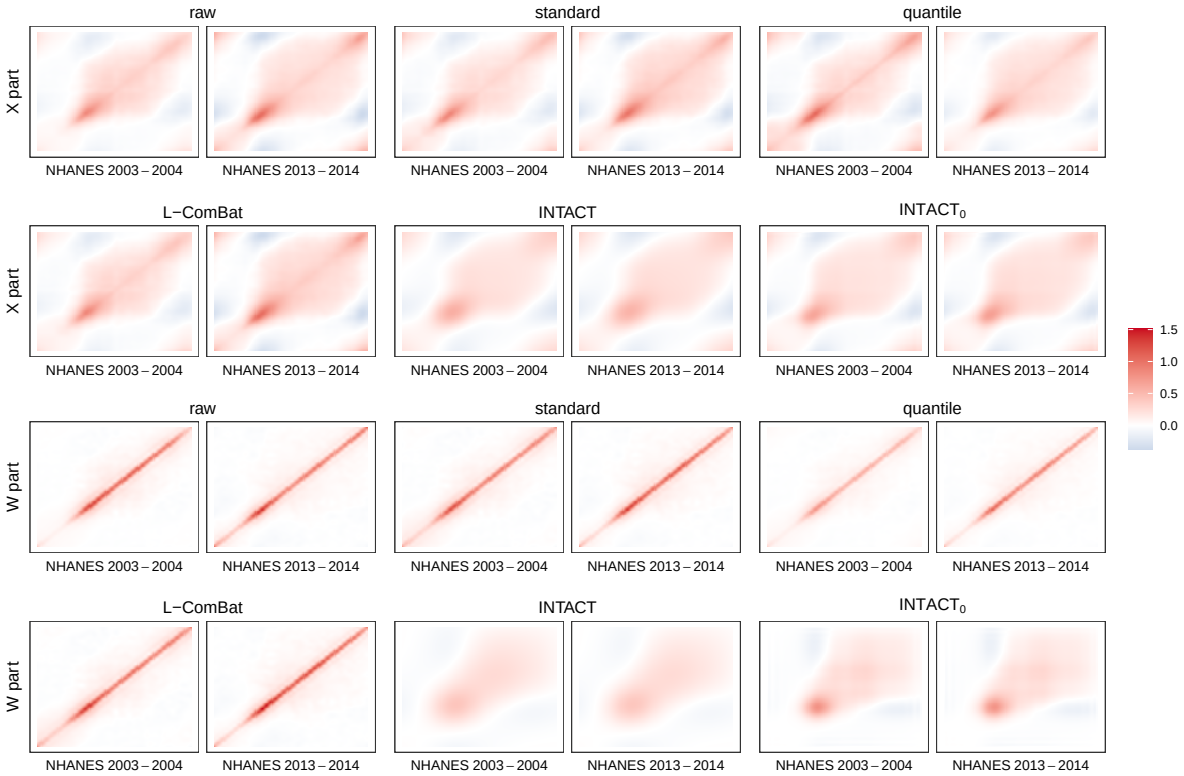

Figure 9: Heatmaps of covariance matrices for raw and harmonized data for the integration of NHANES data.

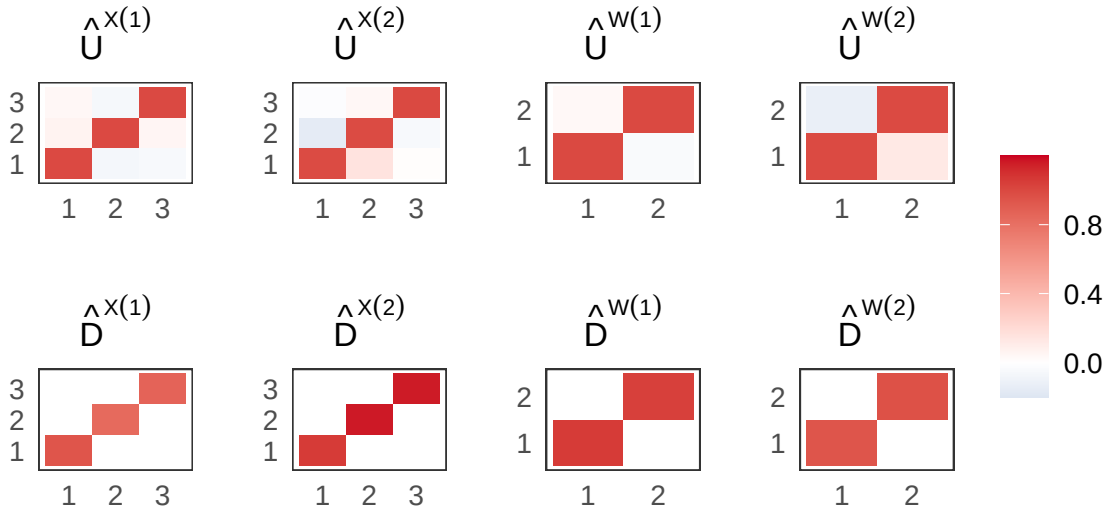

Figure 10: Heatmaps of source-specific rotations and scales for the integration of NHANES data.

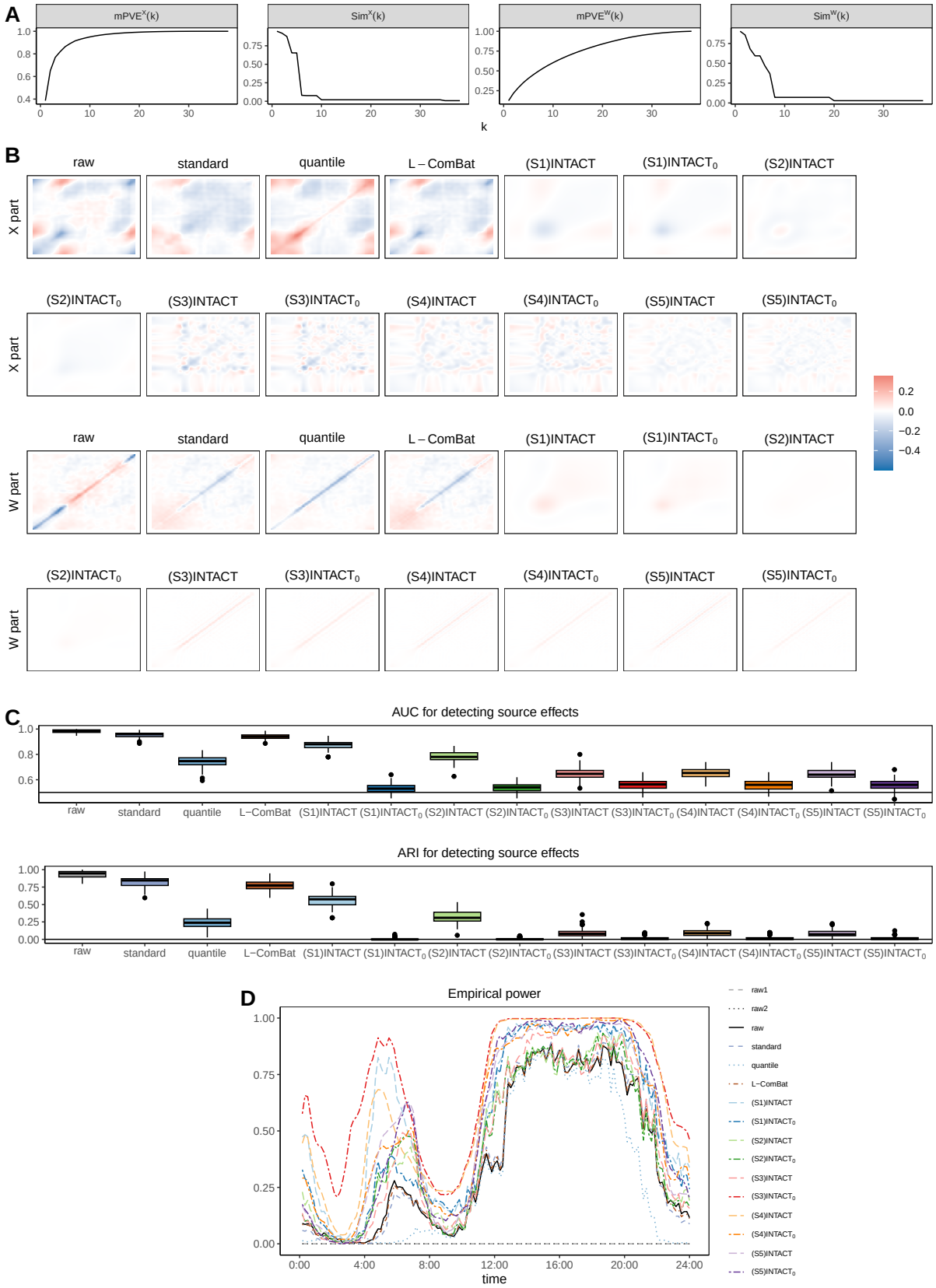

Figure 11: (A–D) Integration results for NHANES physical activity data under different numbers of components. (A) Curves of $\text{mPVE}^X(k)$, $\text{Sim}^X(k)$, $\text{mPVE}^W(k)$, and $\text{Sim}^W(k)$. (B) Heatmaps of covariance differences between the two data sources. (C) Source effect detection. Left: box plots of AUC values; Right: box plots of ARI values. (D) Empirical power for testing the null hypothesis of no age effect.

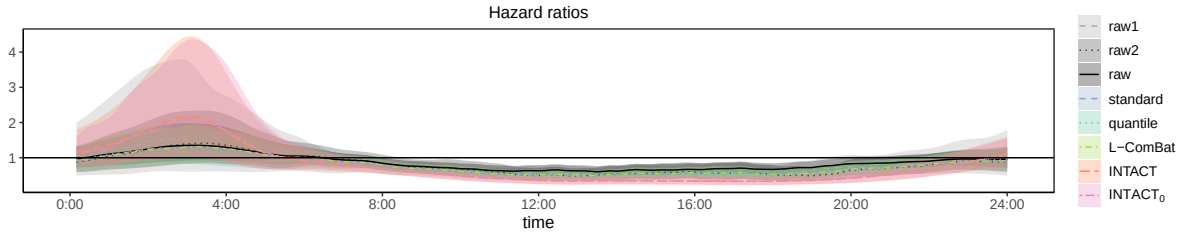

Figure 12: Hazard ratios with Bonferroni-corrected confidence intervals from the Cox model analysis. The black solid line denotes the null value (hazard ratio = 1).

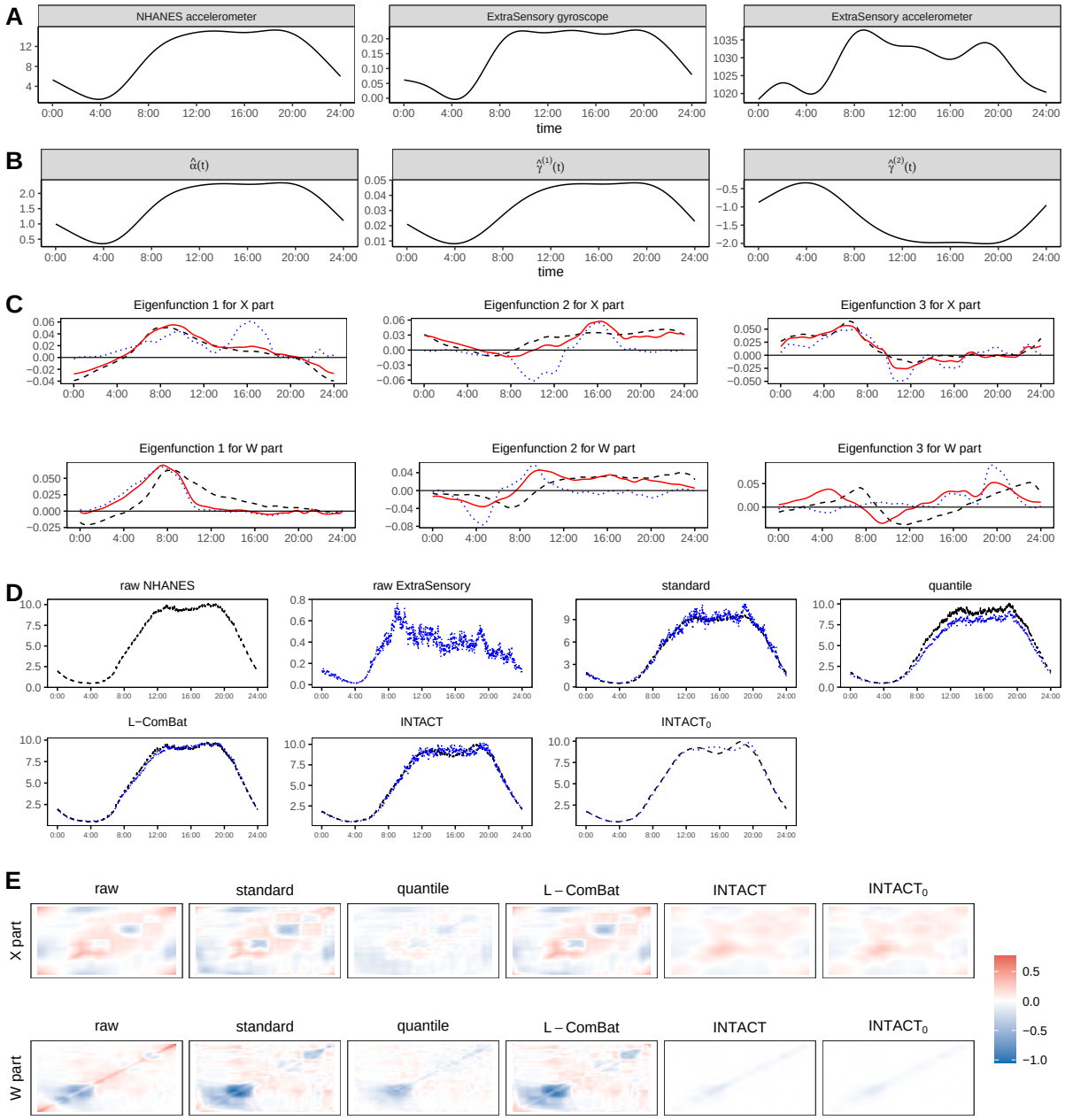

Figure 13: (A) Mean curves for original NHANES accelerometer, ExtraSensory gyroscope, and ExtraSensory accelerometer data. (B–E) Integration results for NHANES accelerometer and ExtraSensory gyroscope data. Abbreviations: raw data (raw), simple standardization (standard), quantile normalization (quantile), longitudinal ComBat (L-ComBat), proposed methods (INTACT, INTACT$_0$). (B) Estimated common mean and source-specific shifts. (C) First three X-part and W-part eigenfunctions. Dashed black: NHANES 2013–2014; dotted blue: ExtraSensory; solid red: common eigenfunctions from INTACT. (D) Mean curves of raw and harmonized data (dashed black: NHANES 2013–2014; dotted blue: ExtraSensory). (E) Heatmaps of covariance differences between sources.