Documentation for Data in "Roadmap2.0Testbed"

1 Background

This dataset is constructed based on a study focusing on Roadmap 2.0, which is a care partner-facing mobile health application [Rozwadowski et al., 2020]. Roadmap 2.0 involves 171 dyads, each consisting of a patient undergoing HCT (target person) and a care partner. Each member of the dyad had the Roadmap mobile app on their smartphone and wore a Fitbit wrist tracker. The Fitbit wrist tracker recorded physical activity, heart rate, and sleep patterns. Furthermore, each participant was asked to self-report their mood via the Roadmap app every evening. The original Roadmap 2.0 data comprises minute-by-minute heart rate records, sleep duration at various timestamps, step count at various timestamps, and daily mood score.

2 Dataset Construction

First, we aggregate the original Roadmap 2.0 data at the daily level. In particular, on day t, for the i-th participant (target person or care partner), we get

$$\begin{aligned}
\text{Heart}_{i,t} &= \text{Average heart rate from 8am on day } t-1 \text{ to 8am on day } t,\\
\text{Sleep}_{i,t} &= \text{Total sleep time from 8pm on day } t-1 \text{ to 8am on day } t,\\
\text{Step}_{i,t} &= \text{Total step count from 8am on day } t-1 \text{ to 8am on day } t.
\end{aligned}$$

Further, we work with the step count at a square root scale, because the distribution of the step count is right skewed: $\text{SqrtStep}_{i,t} = \sqrt{\text{Step}_{i,t}}$. We utilize only the initial 15 weeks of data for each participant, translating to the first 105 days. We exclude dyads that exhibit a missing rate exceeding 75% for any variable within these initial 15 weeks. This filtering process leaves us with 49 dyads. Following the filtering process, we perform missing value imputation: we implement the mice algorithm [Van Buuren and Groothuis-Oudshoorn, 2011] for each individual participant. Furthermore, we consider the average mood of the preceding week; for participant i in week w, we form

$$\text{WeeklyMood}_{i,w} = \text{Average score of mood in week } w-1.$$

We utilize data from Roadmap 2.0 starting from day 8, supplemented by WeeklyMood from the end of week 1 in Roadmap 2.0. Here, day 8 from Roadmap 2.0 will be redefined as day 1, and $\text{WeeklyMood}_{i,2}$ will be recast as $\text{WeeklyMood}_{i,1}$. This results in 14 weeks of data for each participant, translating to 98 days.

We standardize all variables. In particular, to standardize a variable, we use the time specific average and the time specific standard deviation across the same type of individual (target person or care partner). From now on, when we refer to the variables $\text{Heart}_{i,t}, \text{Sleep}_{i,t}, \text{SqrtStep}_{i,t}, \text{WeeklyMood}_{i,w}$, we mean their standardized values.

We model $\text{Heart}_{i,t}, \text{Sleep}_{i,t}, \text{SqrtStep}_{i,t}$ with generalized estimating equations [Hardin and Hilbe, 2012]. In particular, for each participant i (with $c(i)$ being the other person in the dyad) and each daily variable, we fit a separate linear model with the AR(1) working correlation using the GEE approach [Højsgaard et al., 2006]:

$$\begin{aligned}
\text{Heart}_{i,t+1} &\sim \text{Heart}_{i,t} + \text{Sleep}_{i,t} + \text{SqrtStep}_{i,t} + \text{WeeklyMood}_{i,w(t)} + \text{WeeklyMood}_{c(i),w(t)},\\
\text{Sleep}_{i,t+1} &\sim \text{Heart}_{i,t} + \text{Sleep}_{i,t} + \text{SqrtStep}_{i,t} + \text{WeeklyMood}_{i,w(t)} + \text{WeeklyMood}_{c(i),w(t)},\\
\text{SqrtStep}_{i,t+1} &\sim \text{Heart}_{i,t} + \text{Sleep}_{i,t} + \text{SqrtStep}_{i,t} + \text{WeeklyMood}_{i,w(t)} + \text{WeeklyMood}_{c(i),w(t)}\,.
\end{aligned}$$

Here, $w(t)$ is the number of week day t is in. We extract coefficients and residuals from each one of the fitted models. Let β denote the fitted coefficients and ε denote the residuals:

$$\begin{aligned}
\text{Heart}_{i,t+1} &= \left(1, \text{Heart}_{i,t}, \text{Sleep}_{i,t}, \text{SqrtStep}_{i,t}, \text{WeeklyMood}_{i,w(t)}, \text{WeeklyMood}_{c(i),w(t)}\right)^\top \beta_{\text{Heart},i} \\
&\quad + \varepsilon_{\text{Heart},i,t+1}, \\
\text{Sleep}_{i,t+1} &= \left(1, \text{Heart}_{i,t}, \text{Sleep}_{i,t}, \text{SqrtStep}_{i,t}, \text{WeeklyMood}_{i,w(t)}, \text{WeeklyMood}_{c(i),w(t)}\right)^\top \beta_{\text{Sleep},i} \\
&\quad + \varepsilon_{\text{Sleep},i,t+1}, \\
\text{SqrtStep}_{i,t+1} &= \left(1, \text{Heart}_{i,t}, \text{Sleep}_{i,t}, \text{SqrtStep}_{i,t}, \text{WeeklyMood}_{i,w(t)}, \text{WeeklyMood}_{c(i),w(t)}\right)^\top \beta_{\text{SqrtStep},i} \\
&\quad + \varepsilon_{\text{SqrtStep},i,t+1}.
\end{aligned} \tag{1}$$

We model the weekly mood similarly. For each participant i, we fit a separate linear model with the AR(1) working correlation using the GEE approach [Højsgaard et al., 2006] and extract coefficients and residuals:

$$\begin{aligned}
&\text{WeeklyMood}_{i,w+1} \sim \text{WeeklyMood}_{i,w} + \text{WeeklyMood}_{c(i),w}\,. \\
&\text{WeeklyMood}_{i,w+1} = \left(1, \text{WeeklyMood}_{i,w}, \text{WeeklyMood}_{c(i),w}\right)^\top \theta_{\text{WeeklyMood},i} + \eta_{\text{WeeklyMood},i,w+1}.
\end{aligned} \tag{2}$$

3 Data Files

The dataset includes 174 `csv` files. For each pair of target person (i) and care partner (j), the dataset includes the following files.

- `coeffi_pair[pair_no].csv`

  This a 6×3 matrix consisting of $[\beta_{\text{Heart},i}, \beta_{\text{Sleep},i}, \beta_{\text{SqrtStep},i}]$ in (1). These are the fitted coefficients for the daily variables for the target person.

- `coeffi_caregiver_pair[pair_no].csv`

  This is a 6×3 matrix consisting of $[\beta_{\text{Heart},j}, \beta_{\text{Sleep},j}, \beta_{\text{SqrtStep},j}]$ in (1). These are the fitted coefficients for the daily variables for the care partner.

- `residual_pair[pair_no].csv`

  This is a 98×6 matrix consisting of $[\varepsilon_{\text{Heart},j}, \varepsilon_{\text{Sleep},j}, \varepsilon_{\text{SqrtStep},j}, \varepsilon_{\text{Heart},j}, \varepsilon_{\text{Sleep},j}, \varepsilon_{\text{SqrtStep},j}]$ in (1). These are the residuals from the GEE fit for the daily variables.

- `coeffi_weekly_pair[pair_no].csv`

  This is a 3×2 matrix consisting of $[\theta_{\text{WeeklyMood},i}, \theta_{\text{WeeklyMood},j}]$ in (2). These are the fitted coefficients for the weekly mood for both the target person and the care partner.

- `residual_weekly_pair[pair_no].csv`

  This is a 14×2 matrix consisting of $[\eta_{\text{WeeklyMood},i}, \eta_{\text{WeeklyMood},j}]$ in (2). These are the residuals from the GEE fit for the weekly mood.

- `original_pair[pair_no].csv`

  This is a vector of length 8 consisting of

  $$[\text{Heart}_{i,1}, \text{Sleep}_{i,1}, \text{SqrtStep}_{i,1}, \text{Heart}_{j,1}, \text{Sleep}_{j,1}, \text{SqrtStep}_{j,1}, \text{WeeklyMood}_{i,1}, \text{WeeklyMood}_{j,1}].$$

  These are the initial values of the variables.

References

James W Hardin and Joseph M Hilbe. *Generalized estimating equations*. CRC press, 2012.

Søren Højsgaard, Ulrich Halekoh, and Jun Yan. The r package geepack for generalized estimating equations. *Journal of statistical software*, 15:1–11, 2006.

Michelle Rozwadowski, Manasa Dittakavi, Amanda Mazzoli, Afton L Hassett, Thomas Braun, Debra L Barton, Noelle Carlozzi, Srijan Sen, Muneesh Tewari, David A Hanauer, et al. Promoting health and well-being through mobile health technology (roadmap 2.0) in family caregivers and patients undergoing hematopoietic stem cell transplantation: protocol for the development of a mobile randomized controlled trial. *JMIR Research Protocols*, 9(9):e19288, 2020.

Stef Van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, 45:1–67, 2011.