A Bayesian Joint Model for Analysis of the Frequency and Duration of Physical Activity from a Lifestyle Intervention Trial

Juned Siddique[1], Michael J. Daniels[2], Gül İnan[3], Samuel Battalio[1], Bonnie Spring[1], and Donald Hedeker[4]

[1]Department of Preventive Medicine, Northwestern University, Chicago, Illinois, USA

[2]Department of Statistics, University of Florida, Gainesville, Florida, USA

[3]Department of Mathematics, Istanbul Technical University, Istanbul, Turkey

[4]Department of Public Health Sciences, University of Chicago, Chicago, Illinois, USA

Outline

Motivation

Global recommendations on physical activity (PA)

- According to World Health Organization (WHO):
  - Physical activity includes: Leisure time activities (e.g., walking, gardening, or swimming), transportation (e.g., walking or cycling), occupational and household work, games, sports, and so on.
- WHO suggests:
  - "Adults aged 18–64 should do at least 150 minutes a week of moderate-intensity, or 75 minutes a week of vigorous-intensity physical activity."
  - Physical activity should be performed in bouts (episodes) of at least 10 minutes, and preferably, it should be spread throughout the week.

Physical activity and health

- PA has been shown to lower the risk of many diseases such as:
  - Early death,
  - Heart disease,
  - Stroke,
  - Type 2 diabetes, etc.
- But, most adults report no leisure time for physical activity.

Physical activity interventions

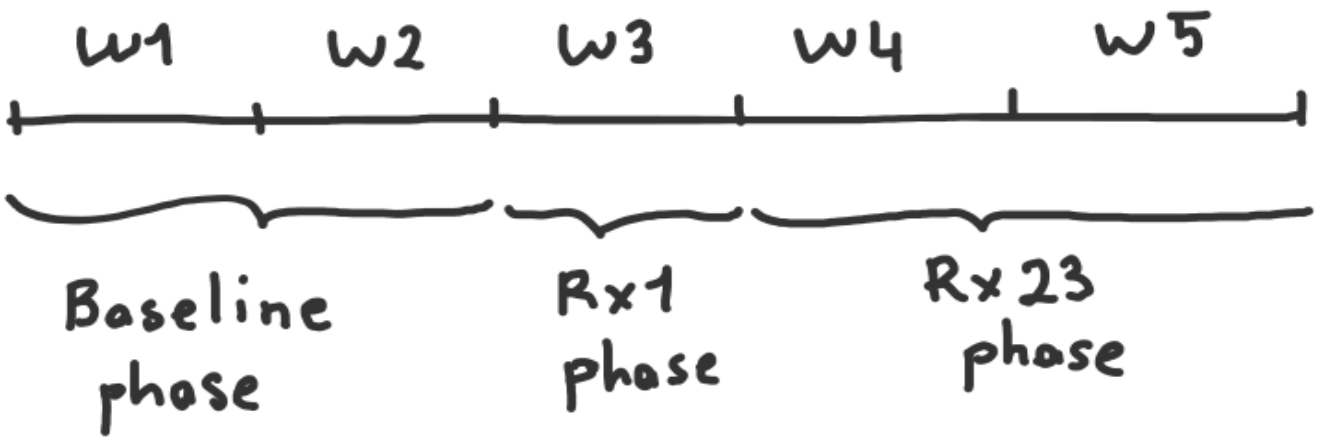
- Behavioral scientists are interested in understanding the frequency and duration of physical activity and how it changes over time.
- Developing effective lifestyle interventions to promote physical activity is still an active area of research.

The Make Better Choices (MBC) study

- MBC was a randomized lifestyle intervention of 204 adults focused on changing PA behaviors of participants (Spring et al., 2010).
- Participants were randomized to one of two activity-related intervention arms:
  - iPA: Increase moderate-to-vigorous PA.
    - The PA goal was 60 min/day of exercise.
  - dSED: Decrease sedentary screen time.
    - The sedentary goal was less than 90 min/day of screen time.

MBC intervention

- A two-week baseline phase.
- A three-week treatment phase:
  - During the first week of treatment (R×1) goal was to attain half of the final target goal and
  - During the last 2 weeks (R×23) goal was to attain the full, targeted goal.



- The participants wore an accelerometer during waking hours for the 5-weeks of the study.
- Thus, it is possible to identify individual PA bouts (episodes) and their duration.

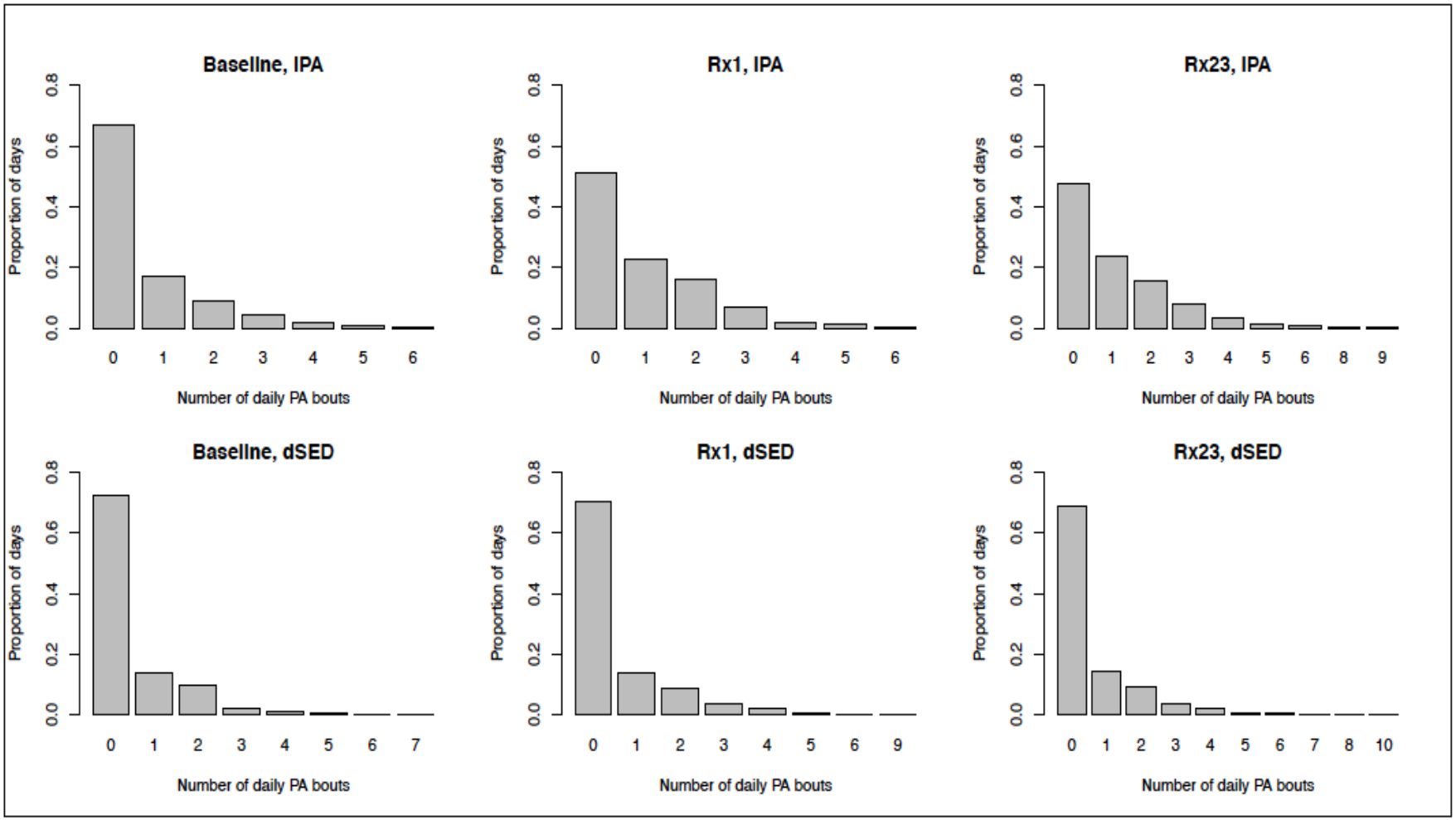


Figure 1: Barplots of number of daily bouts over three different treatment phases (Baseline, Rx1, and Rx23) and two different treatment groups (iPA and dSED) in the MBC study.

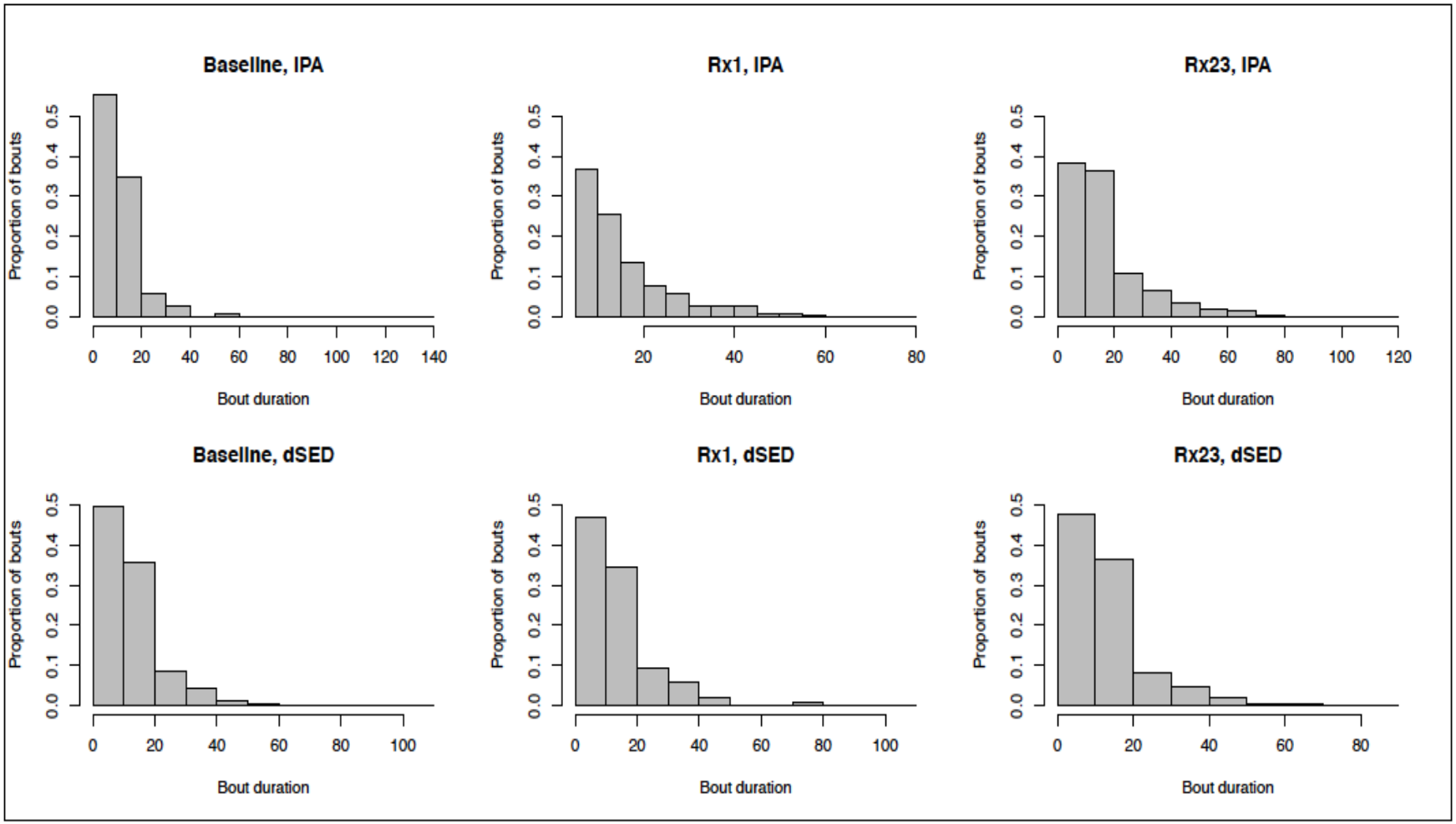


Figure 2: Histogram of bout duration over three different treatment phases (Baseline, Rx1, and Rx23) and two different treatment groups (iPA and dSED) in the MBC study.

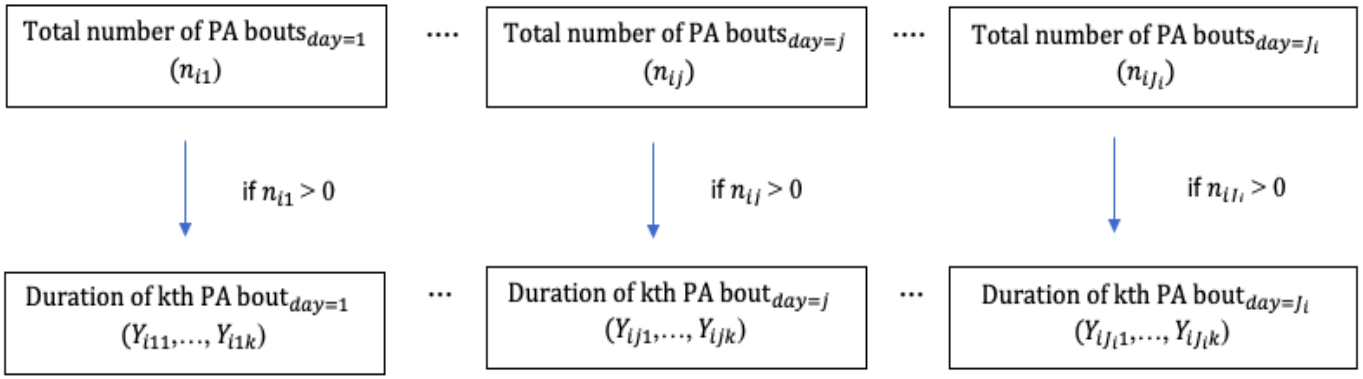
- Figure 1 shows that the proportion of zero-bout days tends to decrease over time in the iPA treatment group and stays constant in the dSED treatment group.
- Figure 1 also shows that participants in the iPA treatment group are more likely to be physically active compared to the those in the dSED treatment group over different treatment phases.
- Figure 1 reveals that the distribution of daily number of bouts is zero-inflated and highly skewed.
- Figure 2 shows that the participants in the iPA treatment group are more likely to have longer episode duration compared to the those in the dSED treatment group over different treatment phases.
- Figure 2 reveals that the duration of bouts is highly skewed.

Aim

- We would like to develop a statistical model that captures unique features of MBC study data:
  - Frequent measurements,
  - Zero-inflated bouts, and
  - Skewed bout durations;
- and allows us to estimate how the number of PA bouts per day and their duration jointly vary over the course of the treatment phase (Baseline, Rx1, and Rx23) and treatment condition (iPA and dSED).

Methodology

Notations



- n_{ij}: The total number of PA bouts for participant i on day j, where $i = 1, \ldots, N$ and $j = 1, \ldots, J_i$.
- Here, N is the total number of participants.
- J_i is the total number of days that the participant i wore an accelerometer throughout the study.
- Y_{ijk}: The duration (in minutes) of kth bout for participant i on (a physically active) day j, where $k = 1, \ldots, n_{ij}$.
- Note that if the participant is physically inactive on day j, then $n_{ij} = 0$, otherwise $n_{ij} > 0$. When $n_{ij} > 0$, Y_{ijk} exists and is always positive.

Distribution assumption for n_{ij}

- Suppose the total number of PA bouts for participant i on day j, n_{ij} results in positive counts ($n_{ij} > 0$), once a hurdle is crossed.
- If the hurdle is not crossed then we observe a count of 0 for n_{ij}.
- Then, we can assume that n_{ij} is generated according to a Bernoulli process such that:

$$n_{ij} = \begin{cases} 0, & \text{with probability } 1 - \pi_{ij}, \\ k, & \text{from a zero-truncated } g(k) \text{ with probability } \pi_{ij}. \end{cases} \quad (1)$$

- where $\pi_{ij} = Pr(n_{ij} > 0)$ is the probability of observing a non-zero PA bout count at day j for the ith participant.
  - Here, the first part provides a zero PA bout-count for physically inactive days, with probability $1 - \pi_{ij}$.
  - The second part provides a non-zero PA bout count ($k > 0$) for physically active days from a stochastic zero-truncated count distribution $g(k)$, with probability π_{ij}.

Distribution assumption for n_{ij}

- Assuming a zero-truncated Poisson (ZTP) distribution for $g(k)$ leads to the following Poisson hurdle distribution:

$$Pr(n_{ij} = k) = \begin{cases} 1 - \pi_{ij}, & k = 0, \\ \pi_{ij} \dfrac{\lambda_{ij}^{k} \exp(-\lambda_{ij})}{k!\left(1-\exp(-\lambda_{ij})\right)} & k > 0. \end{cases} \tag{2}$$

- which is a mixture of a point mass at zero for physically inactive days and a zero-truncated Poisson distribution for physically active days.
- Here λ_{ij} $(\lambda_{ij} > 0)$ is the expected number of PA bouts at day j for the ith participant under an untruncated Poisson dist.
- The model clearly allows for zero-inflation if $(1 - \pi_{ij}) > \exp(-\lambda_{ij})$, which is the probability of observing a zero PA bout count under an untruncated Poisson dist.

Model building for n_{ij}

- We model the parameters in the equation 2 as follows:

$$\begin{aligned} logit(\pi_{ij}) &= \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \mathbf{x}_{1ij}^T\boldsymbol{\beta}_1 + \mathbf{z}_{1ij}^T\mathbf{b}_{1i} \\ \log(\lambda_{ij}) &= \mathbf{x}_{2ij}^T\boldsymbol{\beta}_2 + \mathbf{z}_{2ij}^T\mathbf{b}_{2i}, \end{aligned} \tag{3}$$

- where $\mathbf{x}_{1ij}$ and $\mathbf{x}_{2ij}$ are vector of covariates (e.g., treatment phase and treatment group) for the fixed effects $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively, and
- $\mathbf{z}_{1ij}$ and $\mathbf{z}_{2ij}$ are vector of covariates (e.g., treatment phase and treatment group) for the normally distributed random effects $\mathbf{b}_{1i}$ and $\mathbf{b}_{2i}$, respectively.

Distribution assumption for Y_{ijk}

- For the participants who cross the hurdle at day j, in other words, who have $n_{ij} > 0$, we also assume that duration of kth episode on a physically active day j for the ith participant, Y_{ijk}, follows a mean parametrized Gamma distribution as given below:

$$f(y_{ijk}|n_{ij} > 0) = \frac{1}{\Gamma(\alpha_{ij})}\left(\frac{\alpha_{ij}}{\mu_{ij}}\right)^{\alpha_{ij}} y_{ijk}^{\alpha_{ij}-1} \exp\left(-\frac{y_{ij}\alpha_{ij}}{\mu_{ij}}\right) \quad \text{for } y_{itj} > 0, \quad (4)$$

- where μ_{ij} $(\mu_{ijk} > 0)$ is the expected duration of the kth episode on a physically active day j for the ith participant.
- The shape parameter α_{ij} $(\alpha_{ij} > 0)$ is indexed by participant and day.

Model building for Y_{ijk}

- We model the parameters in the equation 4 as follows:

$$\begin{aligned} \log(\mu_{ij}) &= \mathbf{x}_{3ij}^T \boldsymbol{\beta}_3 + \mathbf{z}_{3ij}^T \mathbf{b}_{3i} \\ \log(\alpha_{ij}) &= \mathbf{x}_{4ij}^T \boldsymbol{\beta}_4 + \mathbf{z}_{4ij}^T \mathbf{b}_{4i}, \end{aligned} \tag{5}$$

- where $\mathbf{x}_{3ij}$ and $\mathbf{x}_{4ij}$ are vector of covariates (e.g., treatment phase and treatment group) for the fixed effects $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$, respectively, and
- $\mathbf{z}_{3ij}$ and $\mathbf{z}_{4ij}$ are vector of covariates (e.g., treatment phase and treatment group) for the normally distributed random effects $\mathbf{b}_{3i}$ and $\mathbf{b}_{4i}$, respectively.

Proposed joint Poisson Hurdle-Gamma model

- The proposed joint Poisson Hurdle-Gamma model is obtained by connecting the models in equations 3 and 5:

$$
\begin{aligned}
logit(\pi_{ij}) &= \mathbf{x}_{ij}^T \boldsymbol{\beta}_1 + \mathbf{z}_{1ij}^T \mathbf{b}_{1i}, \\
\log(\lambda_{ij}) &= \mathbf{x}_{2ij}^T \boldsymbol{\beta}_2 + \mathbf{z}_{2ij}^T \mathbf{b}_{2i}, \\
\log(\mu_{ij}) &= \mathbf{x}_{3ij}^T \boldsymbol{\beta}_3 + \mathbf{z}_{3ij}^T \mathbf{b}_{3i}, \\
\log(\alpha_{ij}) &= \mathbf{x}_{4ij}^T \boldsymbol{\beta}_4 + \mathbf{z}_{4ij}^T \mathbf{b}_{4i}, \quad \text{and} \\
\mathbf{b}_i &= (\mathbf{b}_{1i}, \mathbf{b}_{2i}, \mathbf{b}_{3i}, \mathbf{b}_{4i})^T \overset{i.i.d.}{\sim} MVN(\mathbf{0}, \Sigma),
\end{aligned}
\tag{6}
$$

- where, the first part allows us to associate the probability that a participant will exercise on a given day with study covariates (e.g., treatment phase and treatment group) and
- the second part allows us to model the frequency of bouts on days when a participant exercises.

Proposed joint Poisson Hurdle-Gamma model

- The proposed joint Poisson Hurdle-Gamma model is obtained by connecting the models in equations 3 and 5:

$$
\begin{aligned}
logit(\pi_{ij}) &= \mathbf{x}_{ij}^T \boldsymbol{\beta}_1 + \mathbf{z}_{1ij}^T \mathbf{b}_{1i}, \\
\log(\lambda_{ij}) &= \mathbf{x}_{2ij}^T \boldsymbol{\beta}_2 + \mathbf{z}_{2ij}^T \mathbf{b}_{2i}, \\
\log(\mu_{ij}) &= \mathbf{x}_{3ij}^T \boldsymbol{\beta}_3 + \mathbf{z}_{3ij}^T \mathbf{b}_{3i}, \\
\log(\alpha_{ij}) &= \mathbf{x}_{4ij}^T \boldsymbol{\beta}_4 + \mathbf{z}_{4ij}^T \mathbf{b}_{4i}, \quad \text{and} \\
\mathbf{b}_i &= (\mathbf{b}_{1i}, \mathbf{b}_{2i}, \mathbf{b}_{3i}, \mathbf{b}_{4i})^T \overset{i.i.d.}{\sim} MVN(\mathbf{0}, \Sigma),
\end{aligned}
\tag{6}
$$

- where the third part models the duration of a participant's PA bouts and whether duration changes with respect to study covariates (e.g., treatment phase and treatment group), and
- the fourth part models changes in a participant's bout duration variability.

Proposed joint Poisson Hurdle-Gamma model

- The proposed joint Poisson Hurdle-Gamma model is obtained by connecting the models in equations 3 and 5:

$$
\begin{aligned}
logit(\pi_{ij}) &= \mathbf{x}_{ij}^T \boldsymbol{\beta}_1 + \mathbf{z}_{1ij}^T \mathbf{b}_{1i}, \\
\log(\lambda_{ij}) &= \mathbf{x}_{2ij}^T \boldsymbol{\beta}_2 + \mathbf{z}_{2ij}^T \mathbf{b}_{2i}, \\
\log(\mu_{ij}) &= \mathbf{x}_{3ij}^T \boldsymbol{\beta}_3 + \mathbf{z}_{3ij}^T \mathbf{b}_{3i}, \\
\log(\alpha_{ij}) &= \mathbf{x}_{4ij}^T \boldsymbol{\beta}_4 + \mathbf{z}_{4ij}^T \mathbf{b}_{4i}, \quad \text{and} \\
\mathbf{b}_i &= (\mathbf{b}_{1i}, \mathbf{b}_{2i}, \mathbf{b}_{3i}, \mathbf{b}_{4i})^T \overset{i.i.d.}{\sim} MVN(\mathbf{0}, \Sigma),
\end{aligned}
\tag{6}
$$

- The elements of vector of random effects $\mathbf{b}_i = (\mathbf{b}_{1i}, \mathbf{b}_{2i}, \mathbf{b}_{3i}, \mathbf{b}_{4i})$ represent heterogeneity between the levels of the model for each participant as well as heterogeneity between participants.
- Imposing a multivariate normal distribution on $\mathbf{b}_i$ allows us to connect these four models and to estimate the correlations of these outcomes and their change over time.
- These quantities can help inform reseachers when designing PA interventions.

Further notations & assumptions

- Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4)$ denote the vector of fixed effects in equation 6.
- We assume given $\mathbf{b}_i$, n_{ij} and y_{ijk} are independent within a participant, which is called as conditional independence assumption.
- Let N_i summarize all n_{ij} values for participant i and Y_i summarize their corresponding duration values y_{ijk}.

Marginal likelihood formulation

- The contribution of the ith participant to the marginal likelihood of the data can be calculated as the integration of joint conditional density over the distribution of random effects:

$$
\begin{aligned}
L_i(\boldsymbol{\beta}, \Sigma) &= f(N_i, Y_i|\boldsymbol{\beta}, \Sigma) \\
&= \int_{\mathbf{b}_i} f(N_i, Y_i|\mathbf{b}_i) f(\mathbf{b}_i) d_{\mathbf{b}_i} \\
&= \int_{\mathbf{b}_i} \prod_{j=1}^{N_i} \left[\left\{ (1-\pi_{ij}) \right\}^{I_{(n_{ij}=0)}} \times \left\{ \pi_{ij} \frac{\lambda_{ij}^{n_{ij}} \exp(-\lambda_{ij})}{n_{ij}! \left(1-\exp(-\lambda_{ij})\right)} \right. \right. \\
&\left. \left. \prod_{k=1}^{n_{ij}} \frac{1}{\Gamma(\alpha_{ij})} \left(\frac{\alpha_{ij}}{\mu_{ij}} \right)^{\alpha_{ij}} y_{ijk}^{\alpha_{ij}-1} \exp\left(-\frac{y_{ijk}\alpha_{ij}}{\mu_{ij}} \right) \right\}^{I_{(n_{ij}>0)}} \right] \times \\
&\left(\frac{1}{2\pi} \right)^{\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{b}_i \Sigma^{-1} \mathbf{b}_i) \right\} d_{\mathbf{b}_i},
\end{aligned}
\tag{7}
$$

- where π_{ij}, λ_{ij}, μ_{ij}, and α_{ij} are defined in equation 6.

Data Analysis

The fitted model

- We fitted the following joint model to the MBC study data with the covariates $R1$, $R23$, their interaction with iPA and random effects:

$$
\begin{aligned}
logit(\pi_{ij}) &= \beta_{10} + \beta_{11} R1_i + \beta_{12} R23_i + \beta_{13}(R1_i * iPA_i) + \beta_{14}(R23_i * iPA_i) + \mathbf{b}_{1i}, \\
\log(\lambda_{ij}) &= \beta_{20} + \beta_{21} R1_i + \beta_{22} R23_i + \beta_{23}(R1_i * iPA_i) + \beta_{24}(R23_i * iPA_i) + \mathbf{b}_{2i}, \\
\log(\mu_{ij}) &= \beta_{30} + \beta_{31} R1_i + \beta_{32} R23_i + \beta_{33}(R1_i * iPA_i) + \beta_{34}(R23_i * iPA_i) + \mathbf{b}_{3i}, \\
\log(\alpha_{ij}) &= \beta_{40} + \beta_{41} R1_i + \beta_{42} R23_i + \beta_{43}(R1_i * iPA_i) + \beta_{44}(R23_i * iPA_i) + \mathbf{b}_{4i}, \\
\text{and} \quad \mathbf{b}_i &= (\mathbf{b}_{1i}, \mathbf{b}_{2i}, \mathbf{b}_{3i}, \mathbf{b}_{4i})^T \overset{i.i.d.}{\sim} MVN(\mathbf{0}, \Sigma),
\end{aligned}
$$

- where $i = 1, \ldots, 204$,
- $R1 = 1$ if the treatment phase is $Rx1$, o.w. 0,
- $R23 = 1$ if the treatment phase is $Rx23$, o.w. 0,
- When $R1 = 0$ and $R23 = 0$, then it refers to the baseline.
- $iPA = 1$ if the treatment group is iPA, o.w. 0.

Parameter estimation

- We used a Bayesian approach to estimate the unknown parameters and variables in our model since the optimization of the likelihood function in equation 7 is not tractable.
- The joint posterior distribution of our model is defined as follows:

$$\begin{aligned} p(\boldsymbol{\beta}, \Sigma) &= \text{Marginal Likelihood} \times \text{Prior dist.} \\ &\propto \Big\{ \prod_{i=1}^{N} L_i(\boldsymbol{\beta}, \Sigma) \Big\} \pi(\boldsymbol{\beta}) \pi(\Sigma), \end{aligned} \tag{8}$$

- where we used non-informative prior distributions for model parameters such that:
  - $\pi(\boldsymbol{\beta}) \sim N(0, 1000)$ for each element in $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$, and $\boldsymbol{\beta}_4$.
  - $\pi(\Sigma)$ is the inverse-Wishart distribution having a scale matrix I_p with $(p+1)$ degrees of freedom, where p is the dimension of $\mathbf{b}_i$.

Bayesian settings and posterior checks

- The Markov Chain Monte Carlo (MCMC) algorithm (specifically Gibbs sampling) is used to draw samples from the posterior distribution of the parameters via *JAGS* (version 4.3.0) and the R package *rjags*.
- The first 5000 iterations are considered as burn-in period.
- After burn-in period, 20000 iterations are drawn from 5 different chains in parallel with a thinning value of 20.
- The convergence of the MCMC chains are checked via trace plots, density plots, and Gelman-Rubin statistics provided by R package *coda*.

Table 1: Results from the proposed joint model.

Parameter	**Mean**	95% **CI**
Logistic regression on probability of exercise day		
β_{10} (Int.)	-0.987	[-1.172, -0.800]
β_{11} (R1)	-0.081	[-0.349, 0.185]
β_{12} (R23)	-0.046	[-0.300, 0.207]
β_{13} (R1*iPA)	1.090	[0.728, 1.451]
β_{14} (R23*iPA)	1.031	[0.682, 1.377]
Loglinear regression on mean number of bouts		
β_{20} (Int.)	-0.045	[-0.187, 0.090]
β_{21} (R1)	-0.034	[-0.277, 0.203]
β_{22} (R23)	-0.039	[-0.249, 0.163]
β_{23} (R1*iPA)	0.090	[-0.145, 0.323]
β_{24} (R23*iPA)	0.160	[-0.039, 0.352]

Table 2: Results from the proposed joint model.

Parameter	**Mean**	95% **CI**
Loglinear regression on duration mean		
β_{30} (Int.)	1.540	[1.401, 1.686]
β_{31} (R1)	0.070	[-0.160, 0.294]
β_{32} (R23)	-0.061	[-0.272, 0.144]
β_{33} (R1*iPA)	0.413	[0.176, 0.645]
β_{34} (R23*iPA)	0.639	[0.430, 0.851]
Loglinear regression on duration shape		
β_{40} (Int.)	-0.432	[-0.536, -0.327]
β_{41} (R1)	0.056	[-0.145, 0.268]
β_{42} (R23)	-0.074	[-0.230, 0.083]
β_{43} (R1*iPA)	0.139	[-0.068, 0.343]
β_{44} (R23*iPA)	0.020	[-0.127, 0.171]

Main results on "Logistic regression on probability of exercise day"

- At R×1 treatment phase, the odds of being physically active is 2.9743 ($exp(\hat{\beta}_{13}) = exp(1.090)$) times higher for the iPA treatment group compared to the dSED treatment group.
- At R×23 treatment phase, the odds of being physically active is 2.8039 ($exp(\hat{\beta}_{14}) = exp(1.031)$) times higher for the iPA treatment group compared to the dSED treatment group.
- These results suggest that MBC iPA participants increased their physical activity by having more exercise days.

Main results on "Loglinear regression on mean number of bouts"

- Given that the participant is physically active:
  - The expected number of episodes per day does not change for treatment groups by treatment phase.
- These results also suggest that the number of bouts on days they exercised does not change by treatment phase.

Main results on "Loglinear regression on duration mean"

- Given that the participant is physically active:
  - Furthermore, at R×1 treatment phase, the expected episode duration (in minutes) is increased by 51% ($exp(\hat{\beta}_{33}) = exp(0.413) = 1.5113$) for iPA treatment group compared to the dSED treatment group.
  - At R×23 treatment phase, the expected episode duration (in minutes) is increased by 89% ($exp(\hat{\beta}_{34}) = exp(0.639) = 1.8946$) for iPA treatment group compared to the dSED treatment group.
- These results suggest that MBC iPA participants increased their physical activity by increasing the duration of their exercise bouts.

Conclusion

Conclusion

- Our proposed model can be used to identify features of pyhsical activity and its change over time in terms of frequency of exercise days, number of bouts per day, bout duration, and bout duration variability.
- This will allow behavioural scientists to better understand the effects of their interventions and design new and better interventions.

RESEARCH ARTICLE | Open Access

Joint modeling the frequency and duration of accelerometer-measured physical activity from a lifestyle intervention trial

Juned Siddique, Michael J. Daniels, Gül Inan, Samuel Battalio, Bonnie Spring, Donald Hedeker

First published: 16 September 2023 | https://doi.org/10.1002/sim.9903

References I

Spring, Bonnie et al. (2010). "Make Better Choices (MBC): study design of a randomized controlled trial testing optimal technology-supported change in multiple diet and physical activity risk behaviors". *BMC Public Health* 10.1, p. 586.

Thank you...