Bayesian Inference of Greenspace on Metabolic Equivalence of Task from Observational Actigraphy Experiments

Cindy J. Pang

UCLA Department of Biostatistics
WNAR 2024
X: @CindyJPang
cindyjpang.github.io

June 10, 2024

Physical Activity in the U.S.

- Being physically active is important in mitigating risks associated with preventable diseases including, but not limited to, diabetes, hypertension, and heart disease

Physical Activity in the U.S.

- Being physically active is important in mitigating risks associated with preventable diseases including, but not limited to, diabetes, hypertension, and heart disease
- It is recommended by the Physical Activity Activity Guidelines for Americans that most individuals receive **150-300 minutes of moderate to vigorous physical activity per week**

Physical Activity in the U.S.

- Being physically active is important in mitigating risks associated with preventable diseases including, but not limited to, diabetes, hypertension, and heart disease
- It is recommended by the Physical Activity Activity Guidelines for Americans that most individuals receive **150-300 minutes of moderate to vigorous physical activity per week**
- Despite these recommendations, most Americans **fail to meet these guidelines**

Physical Activity in the U.S.

- Being physically active is important in mitigating risks associated with preventable diseases including, but not limited to, diabetes, hypertension, and heart disease
- It is recommended by the Physical Activity Activity Guidelines for Americans that most individuals receive **150-300 minutes of moderate to vigorous physical activity per week**
- Despite these recommendations, most Americans **fail to meet these guidelines**
- *How do we motivate individuals to be more physically active?*

Greenspace is widely hypothesized to motivate physical activity

As a park and recreation professional, you can influence community health and increase physical activity by providing and promoting safe, equitable, and inclusive access to parks, trails, recreation areas, and green spaces.

People who have more access to green environments, such as parks and trails, tend to walk and be more physically active than those with limited access. The closer people live to a park and the safer they feel in the park, the more likely they are to walk or bike to those places and use the park for physical activity.

Source: CDC.gov

Parks, Recreation and Tourism Management Research

How Parks and Green Spaces Can Improve Your Health

NC State associate professors Lincoln Larson and Aaron Hipp discuss the role of parks and green spaces in promoting well-being.

April 20, 2022 | CNR Web | 4-min. read

However... the evidence is mixed!

conditions related to elevated weight. Literature searches were conducted in SCOPUS, Medline, Embase and PYSCHINFO. Sixty studies met the inclusion criteria and were assessed for methodological quality and strength of the evidence. The majority (68%) of papers found a positive or weak association between greenspace and obesity-related health indicators, but findings were inconsistent and mixed across studies. Several studies found the relationship varied by factors such as age, socioeconomic status and greenspace measure. Developing a theoretical framework which considers the correlates

Lachowycz and Jones, 2011

Results: Access to green space and area levels of crime were not associated with walking for recreation. Distance to facilities had either no or only a small effect on the uptake of different activities. Odds ratios of cycling for leisure dropped as local traffic density increased for both genders. Compared with the lowest quartile for traffic density the likelihood of reporting any cycling for leisure was OR 0.42, (95% CI 0.32 to 0.52, $P < .001$) for women and OR 0.41, (95% CI 0.33 to 0.50, $P < .001$) for men in the highest quartile.

Foster et al., 2009

But, many of these studies are limited by...

- Self-reported measures of physical activity
- Measure "Exposure to Greenspace" using peoples' home, work, or self-reported locations
    - Fails to accurately capture the *dynamic* nature of peoples' daily movements

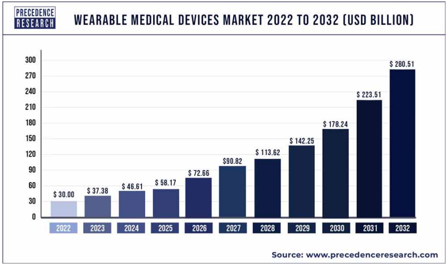
But, many of these studies are limited by...

- Self-reported measures of physical activity
- Measure “Exposure to Greenspace” using peoples’ home, work, or self-reported locations
  - Fails to accurately capture the *dynamic* nature of peoples’ daily movements
  - Lead to underestimating the duration of exposure to greenspace

But, many of these studies are limited by...

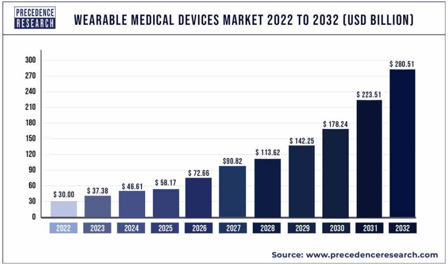
- Self-reported measures of physical activity
- Measure “Exposure to Greenspace” using peoples’ home, work, or self-reported locations
  - Fails to accurately capture the *dynamic* nature of peoples’ daily movements
  - Lead to underestimating the duration of exposure to greenspace
- We can accurately assess exposure with high-resolution measurement devices

The rise of wearable device data (Actigraph Data)



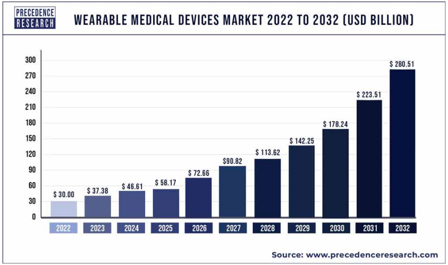
- According to a study of adults in the Health Information National Trends Survey, $\sim 1/3$ Americans uses a wearable device

The rise of wearable device data (Actigraph Data)



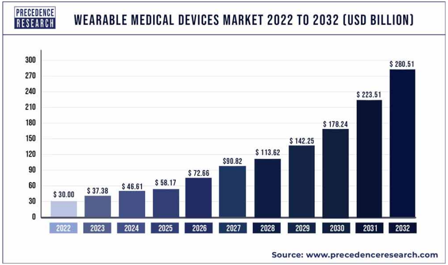
- According to a study of adults in the Health Information National Trends Survey, $\sim 1/3$ Americans uses a wearable device
- Among Wearable Device users, $> 80\%$ would share information with their doctor to support health monitoring

The rise of wearable device data (Actigraph Data)



- According to a study of adults in the Health Information National Trends Survey, $\sim 1/3$ Americans uses a wearable device
- Among Wearable Device users, $> 80\%$ would share information with their doctor to support health monitoring
- People use wearable devices to monitor their daily physical activity levels, sleep cycles, vital signs, and more

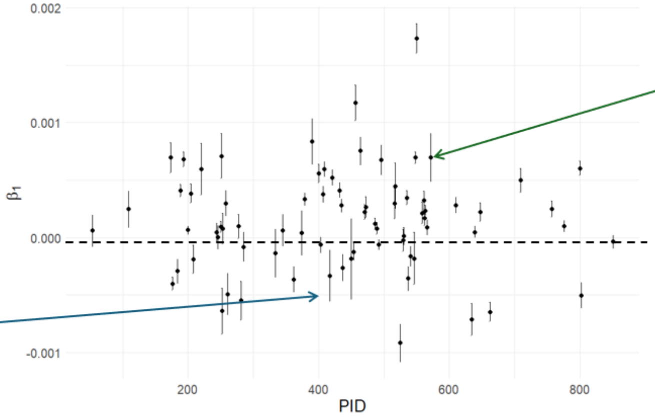
The rise of wearable device data (Actigraph Data)



- According to a study of adults in the Health Information National Trends Survey, $\sim 1/3$ Americans uses a wearable device
- Among Wearable Device users, $> 80\%$ would share information with their doctor to support health monitoring
- People use wearable devices to monitor their daily physical activity levels, sleep cycles, vital signs, and more
- **The data collected by wearable devices can be used to gain insights on how your environment affects your physical activity levels**

Challenges in Analyzing Actigraph Data

1. Variation between Individuals makes it difficult to conclude what the overall effect of an intervention is

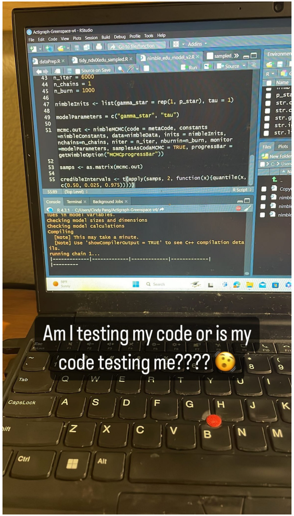


These individuals above the 0-line show that there is a **positive effect** of greenspace on physical activity

... AND these individuals below the 0-line show that there is a **negative effect** of greenspace on physical activity

Challenges in Analyzing Actigraph Data

2. Actigraphy Data are Voluminous and Computationally Intensive



Physical Activity through Sustainable Transport Approaches in Los Angeles (PASTA-LA) Study (Di Loro et al., 2023)

- A cohort study of 460 individuals in the Westwood area monitored for a week in May 2017 and June 2018
- Participants wore Actigraph GT3X+ monitors, GPS devices, and completed online questionaires on demographic data (BMI, Ethnicity, Educational Attainment, Sex, Age)
- Measured Variables:
  - **Normalized Difference Vegetation Index (NDVI)** is a measure of greenness corresponding to a location
  - **Metabolic Equivalence of Task (MET)** measures the intensity of physical activity (i.e. MET < 1.5 is sedentary activity, MET > 6 indicates vigorous exercise)

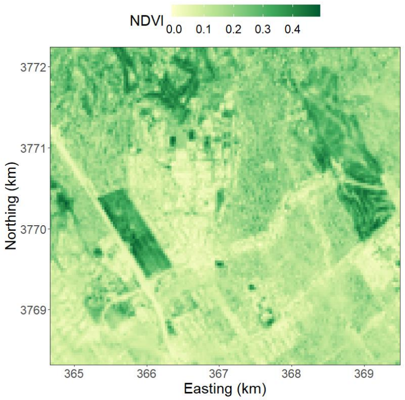
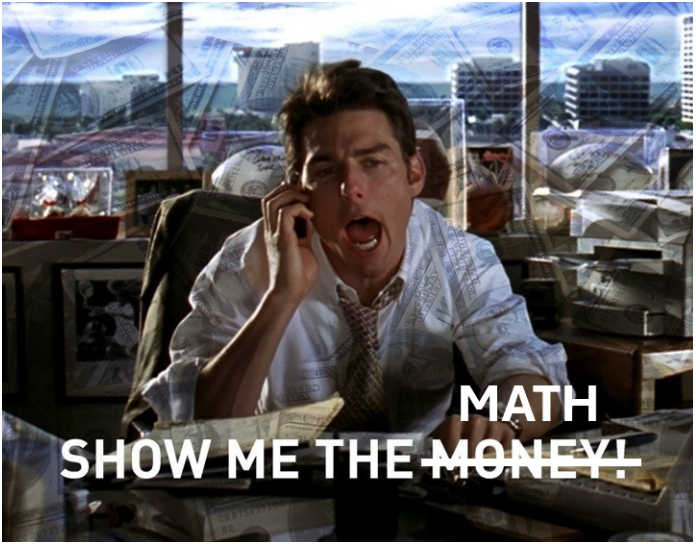


Figure: NDVI in Westwood

Bayesian Subject-Level Pooling or, in other words,



Bayesian Conjugate Model

- Suppose we have M subjects in our study where each subject has n_i observations such that the total number of observations in our study is $N = \sum_{i=1}^{M} n_i$

Bayesian Conjugate Model

- Suppose we have M subjects in our study where each subject has n_i observations such that the total number of observations in our study is $N = \sum_{i=1}^{M} n_i$
- For each realization of the outcome, y_{ij} or MET for subject i observation j, we use the **Bayesian Gaussian Linear Regression Model**:

$$y_{ij} = \beta_{0i} + \beta_{1i} x_{ij} + \epsilon_{ij}; \epsilon_{ij} \overset{ind}{\sim} N(0, \sigma^2) \tag{1}$$

where $x_{ij} =$ NDVI for subject i observation j and $\beta_i = (\beta_{0i}, \beta_{1i})^T$ is a vector containing the slope and intercept for subject i

Bayesian Conjugate Model

- Suppose we have M subjects in our study where each subject has n_i observations such that the total number of observations in our study is $N = \sum_{i=1}^{M} n_i$
- For each realization of the outcome, y_{ij} or MET for subject i observation j, we use the **Bayesian Gaussian Linear Regression Model**:

$$y_{ij} = \beta_{0i} + \beta_{1i} x_{ij} + \epsilon_{ij}; \epsilon_{ij} \overset{ind}{\sim} N(0, \sigma^2) \tag{1}$$

  where x_{ij} = NDVI for subject i observation j and $\beta_i = (\beta_{0i}, \beta_{1i})^T$ is a vector containing the slope and intercept for subject i
- **Priors:**
  - $\beta_i \overset{ind}{\sim} N(\theta, \sigma^2 V_{\beta_i})$ where $V_{\beta_i} = diag(\gamma_{i0}^2, \gamma_{i1}^2)$ for each subject i
  - $\theta \overset{ind}{\sim} N(\mu_\theta, \sigma^2 V_\theta)$ where $\mu_\theta = (\mu_{\theta_0}, \mu_{\theta_1})^T$ and $V_\theta = diag(\delta_0, \delta_1)$
  - $\sigma^2 \overset{ind}{\sim} IG(a_0, b_0)$

Bayesian Conjugate Model

- Suppose we have M subjects in our study where each subject has n_i observations such that the total number of observations in our study is $N = \sum_{i=1}^{M} n_i$
- For each realization of the outcome, y_{ij} or MET for subject i observation j, we use the **Bayesian Gaussian Linear Regression Model**:

$$y_{ij} = \beta_{0i} + \beta_{1i} x_{ij} + \epsilon_{ij}; \epsilon_{ij} \overset{ind}{\sim} N(0, \sigma^2) \tag{1}$$

  where x_{ij} = NDVI for subject i observation j and $\beta_i = (\beta_{0i}, \beta_{1i})^T$ is a vector containing the slope and intercept for subject i
- **Priors:**
  - $\beta_i \overset{ind}{\sim} N(\theta, \sigma^2 V_{\beta_i})$ where $V_{\beta_i} = diag(\gamma_{i0}^2, \gamma_{i1}^2)$ for each subject i
  - $\theta \overset{ind}{\sim} N(\mu_\theta, \sigma^2 V_\theta)$ where $\mu_\theta = (\mu_{\theta_0}, \mu_{\theta_1})^T$ and $V_\theta = diag(\delta_0, \delta_1)$
  - $\sigma^2 \overset{ind}{\sim} IG(a_0, b_0)$
- Our inferential objective is to estimate the subject-wise impact of the NDVI on MET through the β_i's and pooled impact through θ

Augmented System

- Let $Y = (Y_1^T, ..., Y_M^T)^T$ be the $N \times 1$ vector obtained by stacking the n_i vectors Y_i for $i = 1, ..., M$

Augmented System

- Let $Y = (Y_1^T, ..., Y_M^T)^T$ be the $N \times 1$ vector obtained by stacking the n_i vectors Y_i for $i = 1, ..., M$
- Let X_i be a $n_i \times 2$ design matrix for subject i and $X = diag(X_1, ..., X_M)$ is a $N \times 2M$ matrix

Augmented System

- Let $Y = (Y_1^T, ..., Y_M^T)^T$ be the $N \times 1$ vector obtained by stacking the n_i vectors Y_i for $i = 1, ..., M$
- Let X_i be a $n_i \times 2$ design matrix for subject i and $X = diag(X_1, ..., X_M)$ is a $N \times 2M$ matrix
- Collecting these quantities, we have the following linear models:
  - $Y|\beta, \sigma^2 \sim N(X\beta, \sigma^2 I_N)$ where $\beta = (\beta_1, ..., \beta_M)^T$
  - $\beta|\theta, \sigma^2 \sim N(Z\theta, \sigma^2 V_\beta^*)$ where Z is a $2M \times 2$ matrix formed by stacking M I_2 matrices and $V_\beta^* = diag(V_{\beta_1}, ..., V_{\beta_M})$
  - $\theta|\sigma^2 \sim N(\mu_\theta, \sigma^2 V_\theta)$

Augmented System

- Let $Y = (Y_1^T, ..., Y_M^T)^T$ be the $N \times 1$ vector obtained by stacking the n_i vectors Y_i for $i = 1, ..., M$
- Let X_i be a $n_i \times 2$ design matrix for subject i and $X = diag(X_1, ..., X_M)$ is a $N \times 2M$ matrix
- Collecting these quantities, we have the following linear models:
    - $Y|\beta, \sigma^2 \sim N(X\beta, \sigma^2 I_N)$ where $\beta = (\beta_1, ..., \beta_M)^T$
    - $\beta|\theta, \sigma^2 \sim N(Z\theta, \sigma^2 V_\beta^*)$ where Z is a $2M \times 2$ matrix formed by stacking M I_2 matrices and $V_\beta^* = diag(V_{\beta_1}, ..., V_{\beta_M})$
    - $\theta|\sigma^2 \sim N(\mu_\theta, \sigma^2 V_\theta)$
- Rearranging these equations, we obtain the **augmented system**:

$$Y_* = X_* \gamma_* + \epsilon_*; \epsilon_* \sim N(0, \sigma^2 V_*) \tag{2}$$

where $Y_* = (Y, 0, \mu_\theta)^T$, $X_* = \begin{bmatrix} X & 0 \\ -I_{2M} & Z \\ 0 & I_2 \end{bmatrix}$, $\gamma_* = (\beta^T, \theta^T)^T$, and $V_* = diag(I_N, V_\beta^*, V_\theta)$

Posterior Quantities

- We can use the augmented system to obtain the joint posterior density

$$p(\sigma^2, \gamma_* | Y_*) \propto IG(\sigma^2 | a_0, b_0) \times N(Y_* | X_* \gamma_*, \sigma^2 V_*) \tag{3}$$
$$\propto IG(\sigma^2 | a_*, b_*) \times N(Y_* | \hat{\gamma}_*, \sigma^2 (X_*^T V_*^{-1} X_*)^{-1}) \tag{4}$$

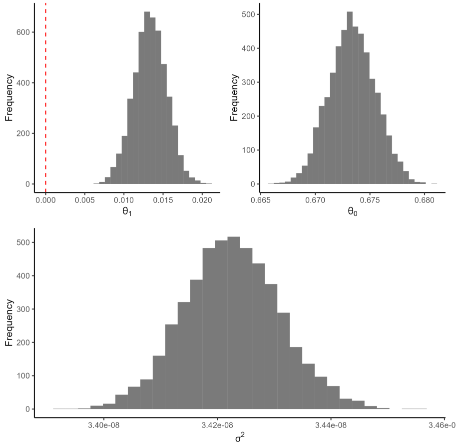
where $a_* = a_0 + \frac{n}{2}$, $b_* = b_0 + \frac{1}{2}(Y - X_* \hat{\gamma}_*)^T V_*^{-1} (Y_* - X_* \hat{\gamma}_*)$ and $\hat{\gamma}_* = (X_*^T V_*^{-1} X_*)^{-1} X_*^T V_*^{-1} Y_*$

- Composition Sampling:
  1. We sample J values $\sigma^2_{(i)} \overset{ind}{\sim} IG(a_*, b_*)$
  2. We draw one value of $\gamma_* \sim N(\hat{\gamma}_*, \sigma^2_{(i)} (X_*^T V_*^{-1} X_*)^{-1})$ J times

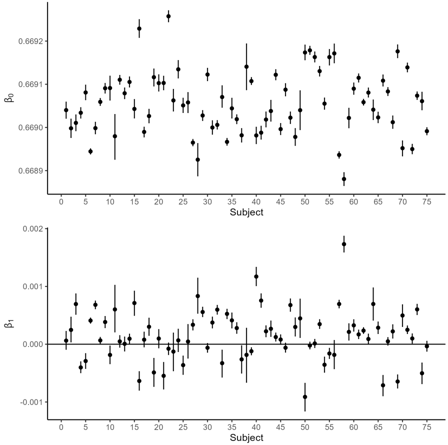
Pooled Parameters $(\theta_0, \theta_1, \sigma^2)$

Table: Parameter Estimates and 95% Credible Intervals for θ_0,θ_1,σ^2

	Quantiles		
Parameter	50%	2.50%	97.50%
θ_0	0.6734	0.6691	0.6776
θ_1	0.0133	0.0091	0.0173
σ^2	3.42E-08	3.41E-08	3.44E-08



Subject-Wise Impact of NDVI on MET (β_i's)



Subject-level point estimates and 95% Credible Intervals for β_0 (top) and β_1 (bottom)

Conclusions

- At the population level, we observe a positive association between green space and physical activity

Conclusions

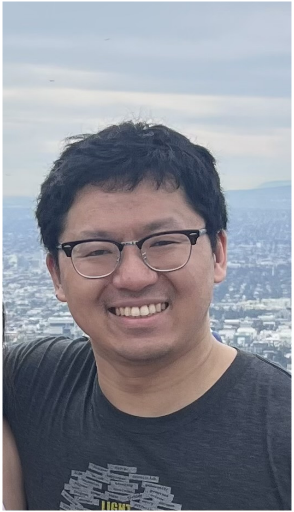
- At the population level, we observe a positive association between green space and physical activity
- The presence of subject-level heterogeneity in the β_1 coefficients suggest that individual characteristics may modulate the impact of greenspace and physical activity levels

Conclusions

- At the population level, we observe a positive association between green space and physical activity
- The presence of subject-level heterogeneity in the β_1 coefficients suggest that individual characteristics may modulate the impact of greenspace and physical activity levels
- The flexibility and scalability of the augmented model makes computation fast and would allow the inclusion of additional covariates and interaction effects

Acknowledgements

Dr. Sudipto Banerjee



Daniel "Dynamic" Zhou

Questions?