Applied and Collaborative Research at CSCC

Beibo Zhao, MS

PhD Candidate

Department of Biostatistics

University of North Carolina at Chapel Hill

beibo@live.unc.edu

UNC Biostatistics / CSCC Faculty Candidate Seminar 2023

October 3, 2023

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

Overview

- Background: CSCC Experience
- Dissertation Research
  - PrecISE: Post hoc subgroup analysis in clinical trials
  - RIVUR: Real data application
- HCHS/SOL Work
  - Temporal patterns of sedentary behavior
  - Retention of Hispanics/Latinos

- **Background: CSCC Experience**
- Dissertation Research
  - PrecISE: Post hoc subgroup analysis in clinical trials
  - RIVUR: Real data application
- HCHS/SOL Work
  - Temporal patterns of sedentary behavior
  - Retention of Hispanics/Latinos

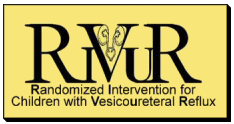
Background: CSCC Experience

Precision Interventions for Severe and/or Exacerbation-Prone Asthma Network (**PrecISE**)

- Developed statistical methods related to subgroup analysis

Hispanic Community Health Study / Study of Latinos (**HCHS/SOL**)

- Student Programmer during MS
- Graduate Research Assistant during PhD
- Developed statistical methods and manuals/guidelines
- Collaborated with investigators on manuscripts



Randomized Intervention for Children with Vesicoureteral Reflux (**RIVUR**)

- Used data as application examples

Overview

- Background: CSCC Experience
- **Dissertation Research**
  - **PrecISE: Post hoc subgroup analysis in clinical trials**
  - **RIVUR: Real data application**
- HCHS/SOL Work
  - Temporal patterns of sedentary behavior
  - Retention of Hispanics/Latinos

Dissertation Research: **PrecISE**

- **PrecISE** is s a multi-arm multi-period crossover randomized clinical trial (RCT) in patients with severe asthma
- Investigating five novel therapies for patients with severe asthma
- Multiple endpoints
  - Forced expiratory volume in one second % predicted
  - 6-item Juniper Asthma Control Questionnaire score
  - CompEx events
- Goals
  - Assess the efficacy of each of the five interventions compared to placebo overall and in a biomarker-defined subgroup
  - Identify subgroups of patients with a large estimated treatment effect based on any of the 3 endpoints (or their combination)

PrecISE
PRECISION INTERVENTIONS FOR SEVERE ASTHMA

(Ivanova et al., 2020)

- Overall efficacy: favorable treatment effect in all patients at the end of RCT
- Discovery of patient subgroups with more pronounced responses than the overall population (perhaps even in absence of an overall effect), and identification of predictive baseline biomarkers
- Guideline-driven: Pre-specifies the subgroups
- Data-driven: Pre-specifies the subgroup selection strategy
    1) Define the targeted subgroups
    2) Estimate the subgroups from data
    3) Confirm the effects in estimated subgroups

(Lipkovich et al., 2023)

- Treatment indicator T (in a RCT, 0 = control, 1 = treatment)
- Candidate biomarkers $\mathbf{X}$
- The treatment effect in a subgroup S

$$\Delta(S) = E[Y^{(1)} - Y^{(0)} \mid \mathbf{X} \in S]$$

- Definitions of the "best" subgroup
  - All individuals with treatment effect > 0 (Imai and Ratkovic, 2013)
  - All individuals with treatment effect > δ (Foster et al., 2011)
- Existing definitions consider only one outcome
- Motivated by PrecISE, which has three endpoints of different types
- Need a definition for the best subgroup to consider
  - Multiple outcomes; specifically, outcomes of different types
  - Both the treatment effect and the size (maximize efficacy)

- For a single outcome, define the best subgroup as the one that maximizes a utility over all possible subgroups (Lai et al., 2014)

$$S_{true} = \arg\max_{S} U(S)$$

- We extend this definition to the case of multiple $J > 1$ outcomes

$$U(S, \boldsymbol{w}) = w_1 U_1(S) + \ldots + w_J U_J(S),$$

where

$$\boldsymbol{w} = (w_1, \ldots, w_J)$$

such that

$$\sum_{j=1}^{J} w_j^2 = 1$$

- Allows for a trade-off between the subgroup size and the treatment effect in the subgroup with respect to each outcome
- Optimal weights are typically not known in advance
- Estimate weights and the best subgroup from data simultaneously

Dissertation Research: **PrecISE**

- Consider continuous, binary, and censored time-to-event outcomes
- Adapt the penalized regression method to estimate best subgroups
- Simulation study
  - Sample size of 400 (200 in each arm)
  - Considered 4 independent uniform biomarkers and three correlated outcomes (continuous, binary, and censored-time-to-event)
  - Compared estimated weights vs. optimal weights
  - Similar estimation accuracy
  - Feasible to estimate weights and subgroup together
- Applicable to a multi-stage randomized clinical trial with prospective enrichment

Dissertation Research: **RIVUR**

- RCT (N = 607) aimed to determine whether long-term antimicrobial prophylaxis is effective, in children with vesicoureteral reflux, to prevent:
  - Primary outcome: Urinary tract infection (UTI) recurrence
  - Secondary outcome: Renal scarring
- Use of long-term antimicrobial prophylaxis treatment may lead to the development of antibiotic resistance and alterations of microbiome
- Number needed to treat (NNT) observed in the trial was large (10)
- Interest in identifying smaller and higher-risk subgroups of children that would benefit the most from long-term treatment

RIVUR
Randomized Intervention for Children with Vesicoureteral Reflux

(Hoberman et al., 2014)

- Estimated best subgroup in the RIVUR trial from 1 or 2 outcomes with the proposed utility-based definition and estimated weights

Outcomes	Size	UTI recurrence		Renal scarring	
		Risk difference (NNT)	p-value	Risk difference (NNT)	p-value
UTI & Renal scarring	85	0.24 (5)	0.02	0.006 (167)	0.95
UTI	427	0.10 (10)	0.01	0.01 (100)	0.72

- Proposed definition allowed incorporating the information from renal scarring into that of UTI recurrence to identify a subgroup with significantly larger clinical benefit (NNT = 5) from long-term antimicrobial prophylaxis than just using UTI recurrence (NNT = 10)

Dissertation Research: **RIVUR** and **PrecISE**

- PrecISE motivated work (submitted)

  Zhao, B., Fine, J., Ivanova, A.(2023). Finding the best subgroup with differential treatment effect with multiple outcomes. Submitted to Statistics in Medicine.

- More RIVUR application (in review)

  Zhao, B., Ivanova, A., Shaikh, N. (2023). Antimicrobial prophylaxis for vesicoureteral reflux: which subgroups of children benefit the most?. Research square, rs.3.rs-3286108. https://doi.org/10.21203/rs.3.rs-3286108/v1

- Subgroup confirmation work (published)

  Zhao, B., Ivanova, A., Fine, J. (2023). Inference on subgroups identified based on a heterogeneous treatment effect in a post hoc analysis of a clinical trial. *Clinical Trials*, *20*(4), 370-379. https://doi.org/10.1177/17407745231173055

Overview

- Background: CSCC Experience
- Dissertation Research
  - PrecISE: Post hoc subgroup analysis in clinical trials
  - RIVUR: Real data application
- **HCHS/SOL Work**
  - **Temporal patterns of sedentary behavior**
  - **Retention of Hispanics/Latinos**

Zhao, B., Sotres-Alvarez, D., Evenson, K.R., Greenlee, H., Mossavar-Rahmani, Y., Qi, Q., Marquez, D.X., Vidot, D.C., Elfassy, T., et al., 9900. **Day-of-the-Week and Time-of-the-Day Patterns of Sedentary Behavior** in the Hispanic Community Health Study / Study of Latinos. Medicine & Science in Sports & Exercise.

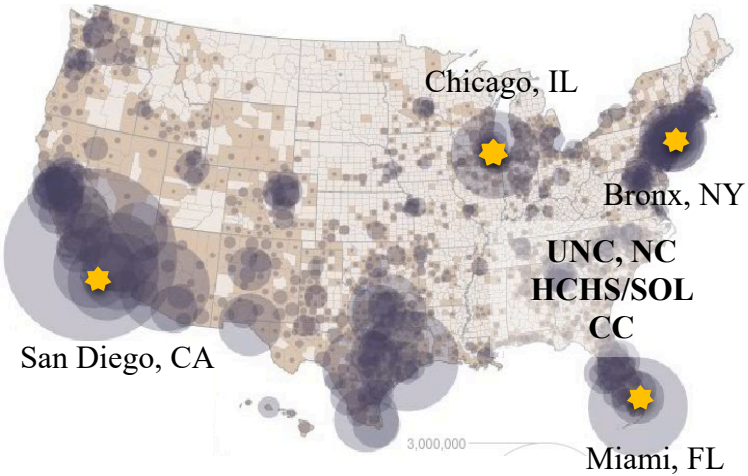
Moderated poster abstract presentation, AHA EPI 2023.

Perreira, K.M., Abreu, M.D.L.A., **Zhao, B.**, Youngblood, M.E., Alvarado, C., Cobo, N., Crespo-Figueroa, M., Garcia, M.L., Giachello, A.L., et al., 2020. **Retaining Hispanics**: Lessons From the Hispanic Community Health Study/Study of Latinos. American Journal of Epidemiology 189:518-31.

"**Multilevel Modelling** for Analyzing Correlated Data with Complex Sampling Design in HCHS/SOL". https://sites.cscc.unc.edu/hchs/node/12075

HCHS/SOL **Analysis Methods – Visit 2**. https://sites.cscc.unc.edu/hchs/node/6113

- **HCHS/SOL** is a multi-center, community-based cohort study (N = 16,415 US Hispanic/Latino adults, 18-74 yrs) at four urban field centers (Baseline, 2008-2011)
- Selected through a stratified multi-stage area probability sample design (Lavange et al., 2010)



Pew Hispanic Center (www.pewhispanic.org)

- Time spent in sedentary behavior is associated with incident cardiovascular disease (CVD)
- Conventional interventions to reduce sedentary behavior with a full-day approach have achieved mixed results
- An alternative strategy may be to target specific periods during the day and/or the week
- The goal is to examine **day-of-the-week** and **time-of-the-day** temporal patterns of sedentary behavior among U.S. Hispanic/Latino adults, overall and by sociodemographic characteristics

- Physical Activity
  - 1-week accelerometer (counts/minute)
  - Non-wear time determined by Choi algorithm (Choi et al., 2011)
  - Adherence: ≥ 3 days with ≥ 10 hrs of wear time per day
- Participants were instructed to remove the accelerometer device during water activities and sleep periods, but they might not do so
- Accelerometer cannot differentiate sleep from sedentary behavior

- **Sedentary bout**
  - An interval of one or more consecutive minutes in which the accelerometer registered <100 counts/min (sedentary time)
- **Sedentary break**
  - An interval of one or more consecutive minutes in which the accelerometer registered ≥ 100 counts/min (active time), between two sedentary bouts
- Non-wear time were not included in any bout or break
- Measures of sedentary behavior:
  - Total volume of sedentary time (mins/day)
  - Total volume of time in sedentary bouts ≥ 60 mins (mins/day)
  - Total number of sedentary breaks (/day)

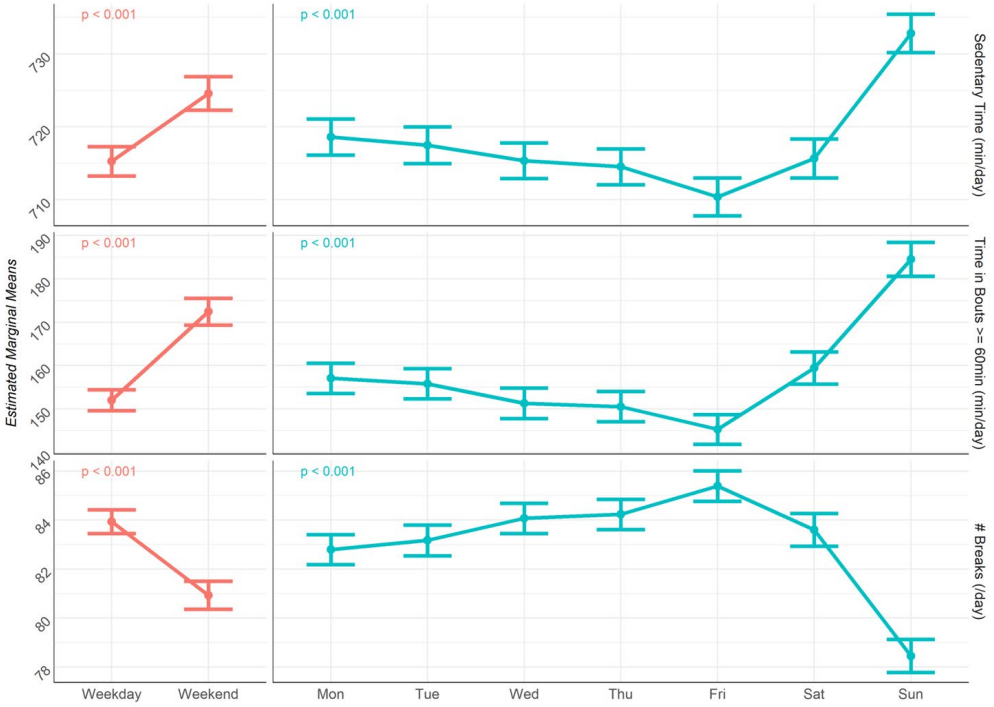
- **Day-of-the-week**
  - Analyzed by days of week
  - Summarized into weekday/weekend
- **Time-of-the-day**
  - Analyzed by 3-hr periods
  - Summarized into 6-hr periods
- Only considered data from adherent days (≥ 10 hrs of wear)
- Excluded excessive wear time (> 23hrs) and device malfunction (Moore et al., 2023)
- Final analytic sample size N = 12,241

- Multivariate-adjusted multi-level mixed-effects linear regression
  - Adjusted for sex, age group, field center, Hispanic/Latino background, employment, shift schedule, BMI (WHO), born in US, acculturation, education, income, season, moderate to vigorous physical activity, Short-Form 12 Health Survey mental and physical score, and wear time
  - Survey designs
    - Multi-level stratified and weighted pseudolikelihood
    - PSU-level (block groups) stratification
    - Actical multi-level inverse probability weighting (IPW) weights, accounting for complex survey design and missing Actical data
  - Random intercepts (households, participants) for clustering effect
  - Interaction between wear time and cross-classification of field center and Hispanic/Latino background

- Tested the significant interaction between time periods and each of the nine subgroups in separate models respectively
- Stratified analysis
  - By age, employment-shift, acculturation, field center, season

- Sensitivity analyses
  - Participants with at least one weekend day (N = 10,837)
  - Removing the 0:00 to 6:00 time period (N = 12,241)
  - Simultaneously account for weekend non-wear and wear during sleep, including only participants with at least one weekend and removing data from the 0:00 – 6:00 time period (N = 10,837)

- Sedentary behavior metrics by day of the week



- Sedentary behavior metrics by time of the day

- Discussion
  - Statistically significant variations in temporal patterns across day-of-the-week and time-of-the-day time periods for all three metrics
  - More sedentary on weekends than on weekdays, and most sedentary on Sundays
  - U-curve pattern, most sedentary late at night, less sedentary throughout the day, reached peak physical activity around noon, gradually more sedentary into the evening
  - Patterns were largely robust across seasonality and most sociodemographic characteristics

- Conclusion
  - Early mornings, evenings, weekends - the more sedentary periods- present windows of opportunity to reduce sedentary time

HCHS/SOL

- Examined the retention of Hispanics/Latinos in HCHS/SOL
- Across 5 years (2009 – 2016), HCHS/SOL maintained high contact, response, and participation rates
- The most difficult Hispanic/Latino populations to retain included young, single, US-born males with less than a high school education
- HCHS/SOL participants primarily sought to help their community and learn more about their health

Field Center and Response Status	Follow-up Interview											
	AFU1		AFU2		AFU3		AFU4		AFU5		Overall (AFUs 1–5)	
	No. of Persons	%	No. of Persons	%	No. of Persons	%	No. of Persons	%	No. of Persons	%	No. of Persons	%
Overall	16,197		16,135		16,067		16,009		15,930		15,930	
Contacted		96.5		95.0		93.4		91.3		89.8		83.3
Responded		95.5		94.3		92.7		90.3		88.6		81.3
Participated		87.9		87.2		84.4		82.2		81.1		67.2

- HCHS/SOL Analysis Methods - Visit 2
  - Provided data analytic guidelines to all HCHS/SOL investigators
  - Created multilevel sampling weights
  - Updated complex survey procedures and model-based procedures
  - Added multilevel modelling section
  - Contributed heavily to survival analysis section
  - Provided examples codes in 4 programming languages (SAS/SUDAAN, R, Stata, Mplus)
- Ongoing
  - Interval-censored survival analysis
  - Inverse sampling
  - Visit 3 variable definitions
  - …

Acknowledgement

- SPECIAL thanks to PrecISE Study PI, and my Dissertation Adviser: Anastasia Ivanova

- All co-authors
- HCHS/SOL Supervisor: Daniela Sotres-Alvarez
- HCHS/SOL Study PI: Jianwen Cai
- Staff and participants of HCHS/SOL

- Funding from National Heart, Lung, and Blood Institute (NHLBI)

References

Ivanova, A., Israel, E., Lavange, L.M., Peters, M.C., Denlinger, L.C., Moore, W.C., Bacharier, L.B., Marquis, M.A., Gotman, N.M., et al., 2020. The precision interventions for severe and/or exacerbation-prone asthma (PrecISE) adaptive platform trial: statistical considerations. Journal of Biopharmaceutical Statistics 30:1026-37.

Lipkovich, I., Svensson, D., Ratitch, B., Dmitrienko, A., 2023. Overview of modern approaches for identifying and evaluating heterogeneous treatment effects from clinical data. Clinical Trials 20:380-93.

Lipkovich, I., Dmitrienko, A., & B. R. D' Agostino, S. (2017). Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. *Stat Med*, *36*(1), 136-196. https://doi.org/10.1002/sim.7064

Imai, K., Ratkovic, M., 2013. Estimating treatment effect heterogeneity in randomized program evaluation. The Annals of Applied Statistics 7:443-70, 28.

Foster, J.C., Taylor, J.M.G., Ruberg, S.J., 2011. Subgroup identification from randomized clinical trial data. Statistics in Medicine 30:2867-80.

Lai, T.L., Lavori, P.W., Liao, O.Y.-W., 2014. Adaptive choice of patient subgroup for comparing two treatments. Contemporary Clinical Trials 39:191-200.

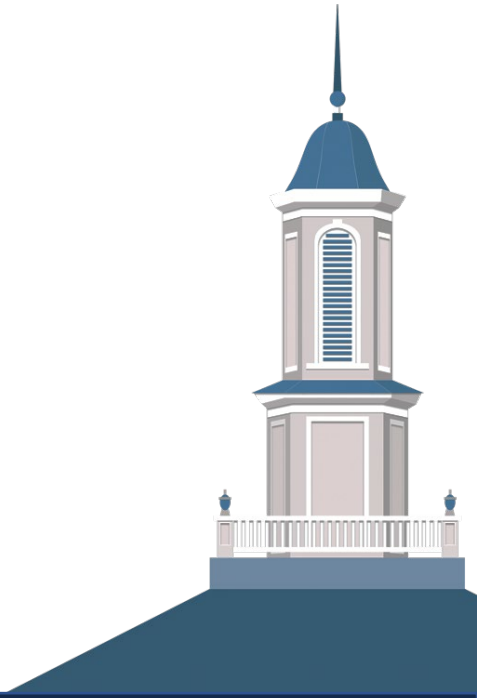
Zhang, Z., Li, M., Lin, M., Soon, G., Greene, T., Shen, C., 2017. Subgroup selection in adaptive signature designs of confirmatory clinical trials. Journal of the Royal Statistical Society. Series C (Applied Statistics) 66:345-61.

Lavange, L.M., Kalsbeek, W.D., Sorlie, P.D., Avilés-Santa, L.M., Kaplan, R.C., Barnhart, J., Liu, K., Giachello, A., Lee, D.J., et al., 2010. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. Ann Epidemiol 20:642-9.

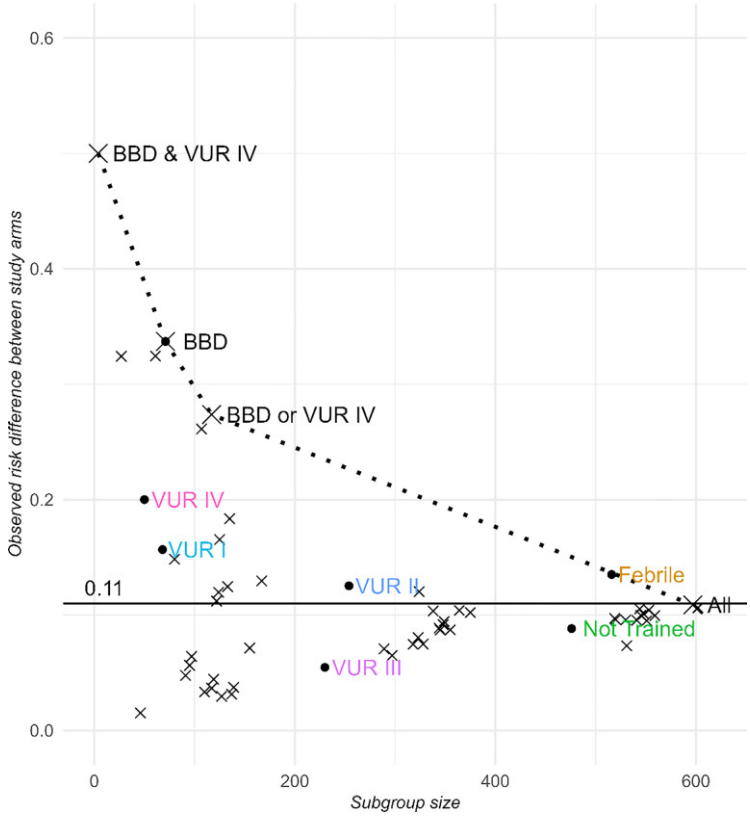
Choi, L., Liu, Z., Matthews, C.E., Buchowski, M.S., 2011. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 43:357-64.

Moore, C.C., Cuthbertson, C.C., Sotres-Alvarez, D., Castaneda, S.F., Cordero, C., Daviglus, M.L., Mossavar-Rahmani, Y., Perreira, K.M., Evenson, K.R., 2023. Step-based metrics and translations of physical activity guidelines among adults in the HCHS/SOL. Medicine & Science in Sports & Exercise 55.

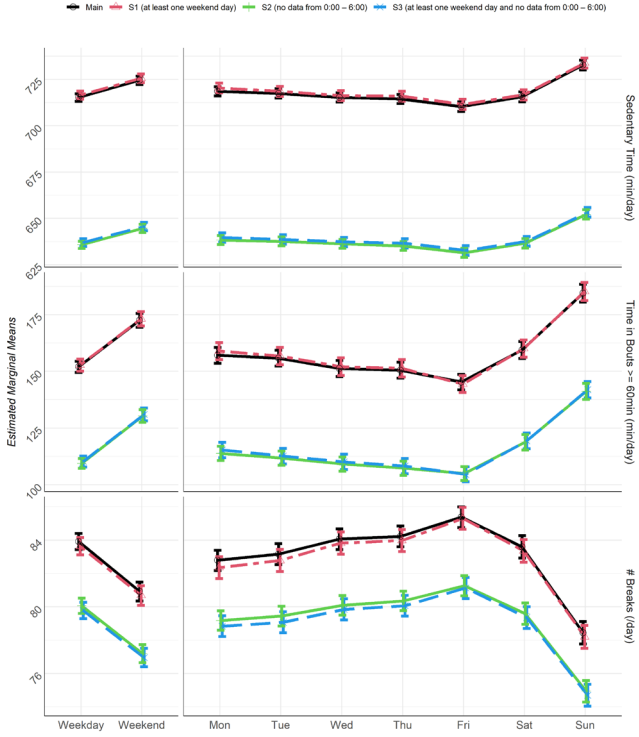
Thank You

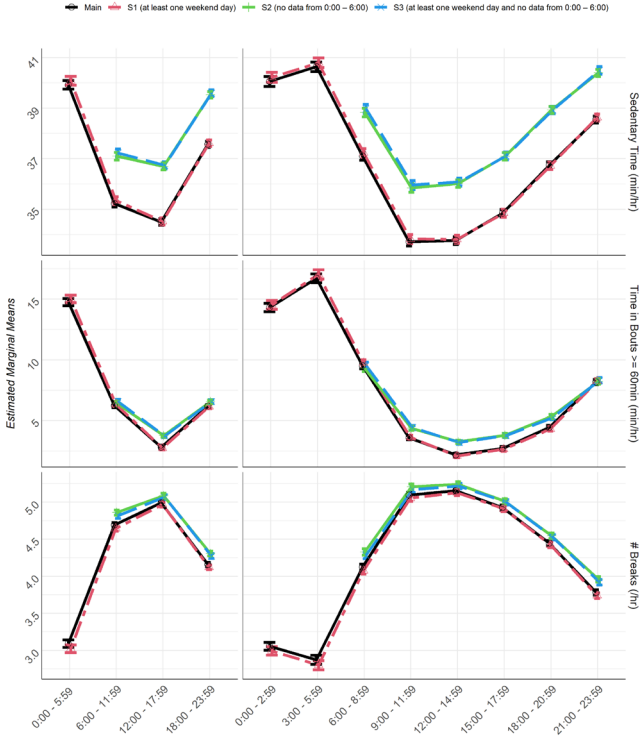


More RIVUR


0.6
0.4
0.2
0.0
Observed risk difference between study arms
BBD & VUR IV
BBD
BBD or VUR IV
VUR IV
VUR I
0.11
VUR II
Febrile
All
Not Trained
VUR III
0
200
400
600
Subgroup size

- Sensitivity Analyses





Collaborative Research at CSCC: HCHS/SOL

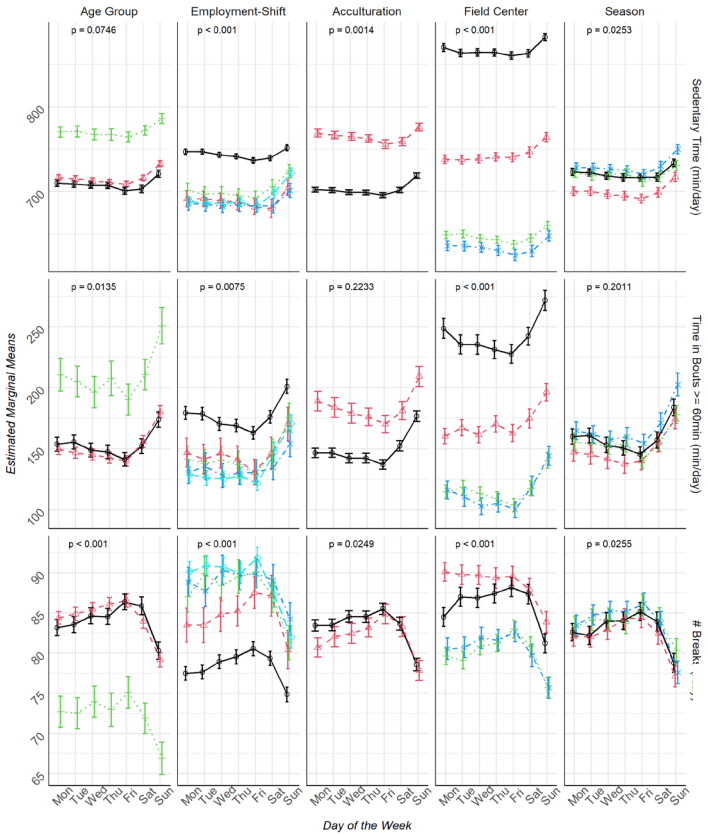


Figure 3. Estimated marginal means per day (95% CI) in sedentary behavior metrics by day of the week, stratified by subgroups, HCHS/SOL 2008-2011 (N = 12,241)

- Day-shift workers exhibited a sizeable uptick in sedentary time from Friday to Saturday while others did not show much change
- Chicago site exhibited an opposite trend for all sedentary metrics during weekdays comparing to other field centers

Collaborative Research at CSCC: HCHS/SOL

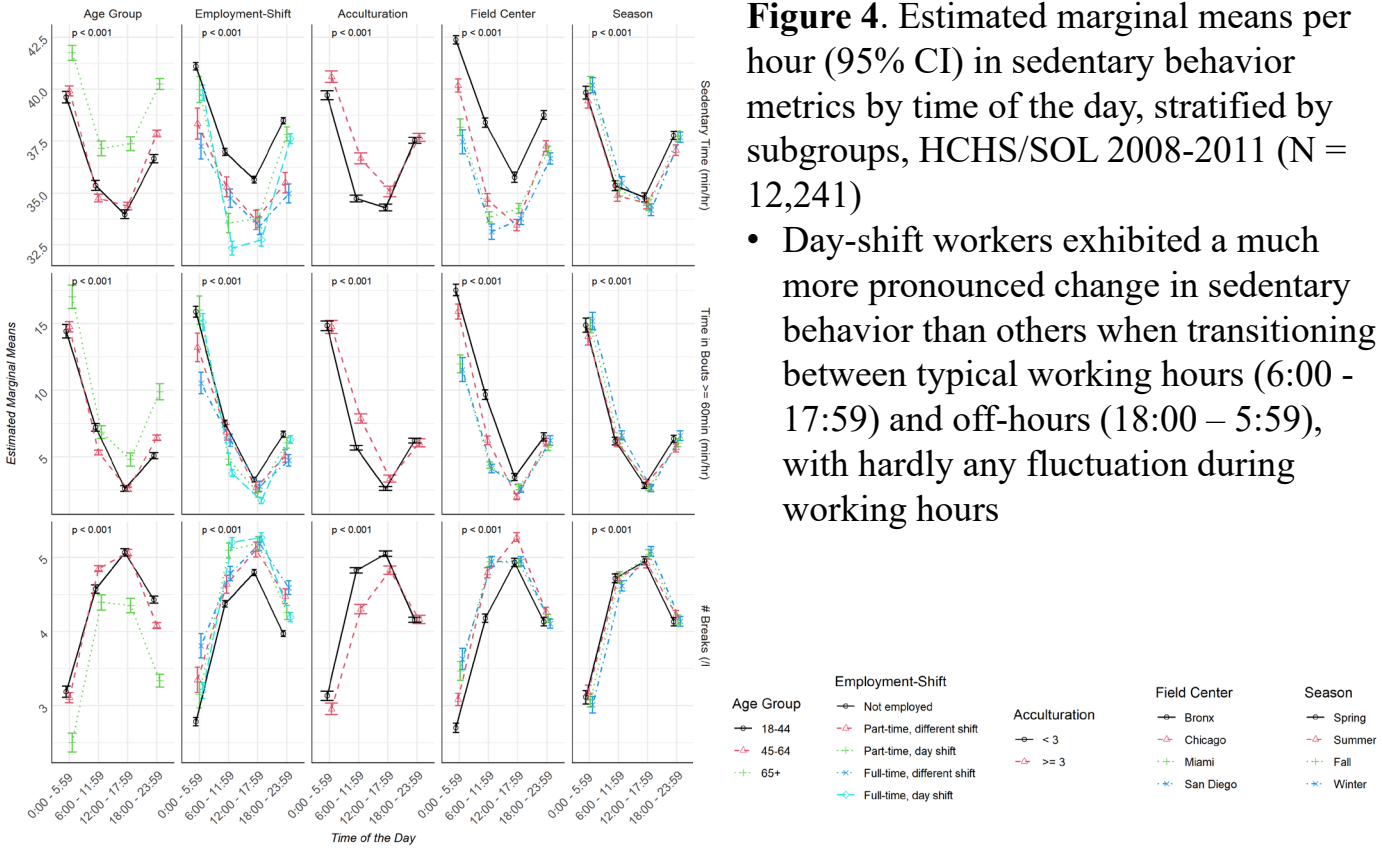


Figure 4. Estimated marginal means per hour (95% CI) in sedentary behavior metrics by time of the day, stratified by subgroups, HCHS/SOL 2008-2011 (N = 12,241)

- Day-shift workers exhibited a much more pronounced change in sedentary behavior than others when transitioning between typical working hours (6:00 - 17:59) and off-hours (18:00 – 5:59), with hardly any fluctuation during working hours