Quantifying Physical Activity Characteristics with Accelerometry Data: Application to the National Health and Nutrition Examination Survey (NHANES)
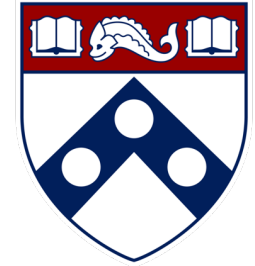
Amelia H. Tran

December 6, 2022

1

Background

Overview

➢ **Goals:**

    ➢ Process accelerometry National Health and Nutrition Examination Survey (NHANES) data from CDC

    ➢ Investigate sample NHANES to create a pipeline for clinical development

➢ **Dataset:** minute-level accelerometry data from NHANES

➢ **Approach:**

    ➢ Process minute-level accelerometry NHANES data:

        ➢ Parse data into <u>subject-level datasets</u>

        ➢ Obtain <u>minute-level timestamps</u> when measurements were taken

        ➢ Quantify <u>minute-level physical activity characteristics</u> with an open-source package

    ➢ Characterize physical actigraphy, demographic, and medical characteristics of NHANES dataset

➢ **Primary Findings:**

    ➢ Quantified subject-level and day-level physical activity characteristics with ***Arctools*** package

    ➢ Generated exploratory visualizations and statistical analysis of NHANES characteristics and physical activity

Agenda

Accelerometry Data
- Overview of data from wearable devices
- Key challenges and opportunities

Process Data
- Quantification of physical activity characteristics with *Arctools*
- Summary statistics and graphics of physical activity

NHANES
- Overview of NHANES
- Quantification minute-level accelerometry data from NHANES

Analysis
- Characterization of NHANES with respect to physical activity
- Exploratory visualizations and statistical analysis

Conclusion
- Limitations of current approach
- Future research

# Accelerometry data present challenges and opportunities

➢ **Objective measurements** of physical activity in both controlled and free-living environment

➢ Accelerometer measures **acceleration** amplitude relative to the Earth gravity

➢ **Challenges** with data heterogeneity**:**
  ➢ Sensor location: hip, wrist, ear, thigh, etc.
  ➢ Device rotation
  ➢ Sampling frequency
  ➢ Device calibration

➢ **Opportunities** in health research:
  ➢ Digital biomarker development
  ➢ Pattern of human movement
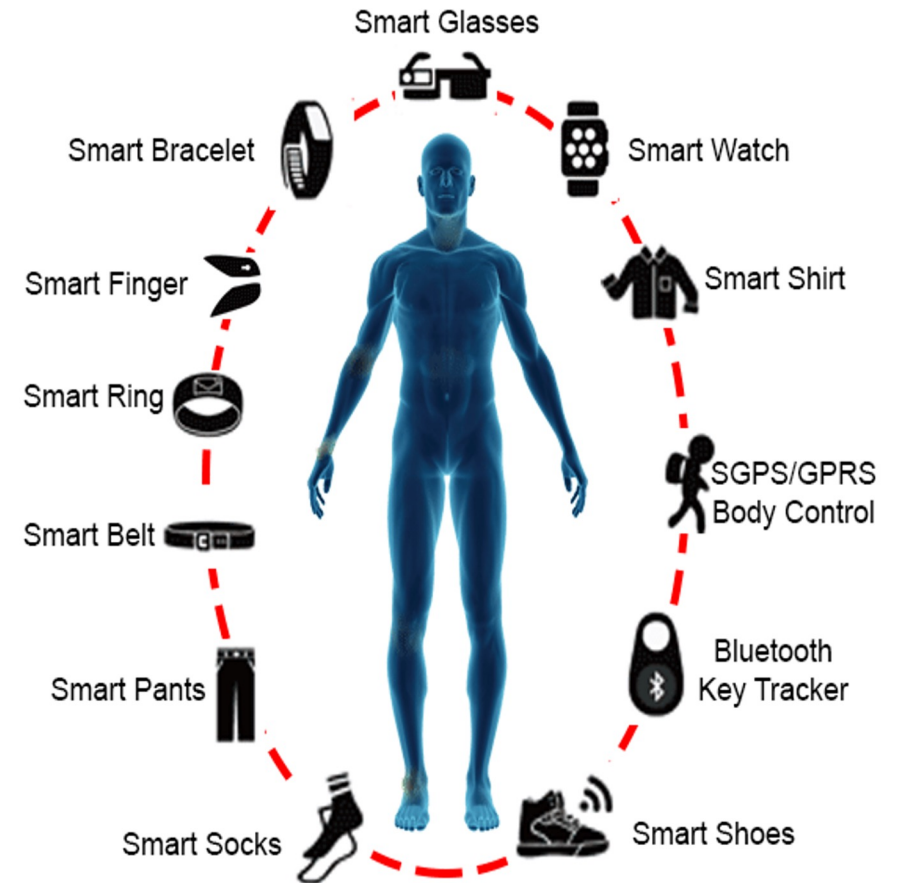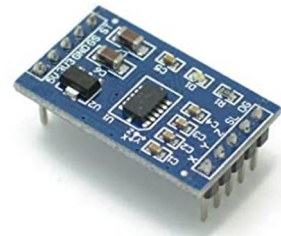
Rodrigues, Joel JPC, et al. "Enabling technologies for the internet of health things." *IEEE Access* 6 (2018): 13129-13141.

Raw accelerometry data is collected in a 3D time series format

➤ Raw data collected along **three orthogonal axes** XYZ

➤ **Universal** acceleration measurements comparable across devices

➤ Output is **three-dimensional time series accelerations** in gravitational units (1g = 9.8 m/s^2)

➤ Device's frame of reference: up-down, left-right, backward-downward
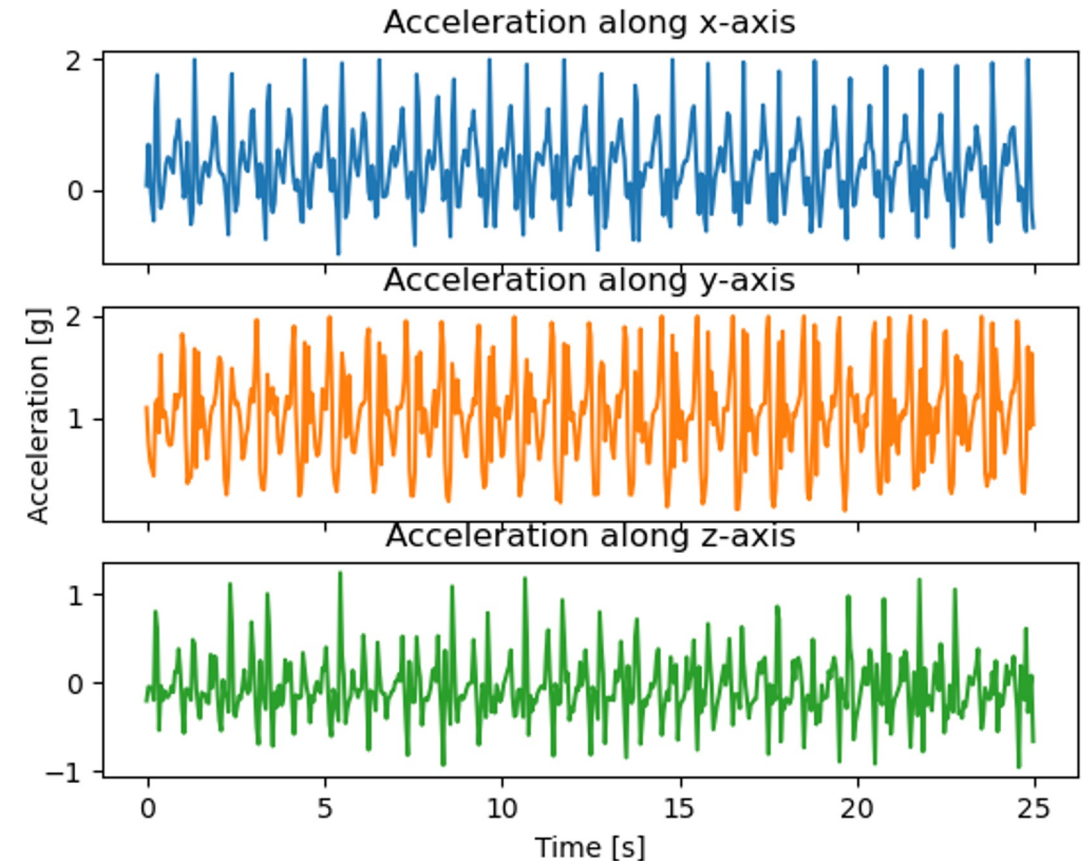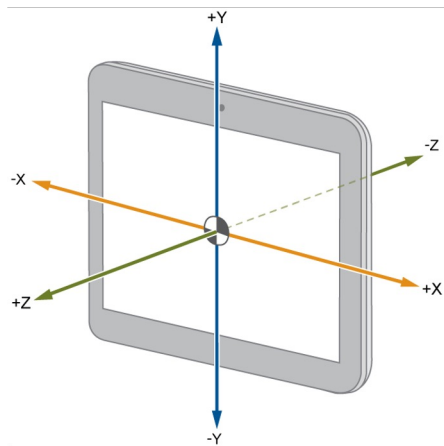
Peter Hausamann, "Activity recognition from accelerometer data" (2020)

Raw data can be collapsed into minute-level data

➢ Acceleration times series $\{x(t), y(t), z(t)\}$ sampled at frequency f

➢ Raw data from XYZ tri-axes are **collapsed** into one value, called **minute-level**

➢ **Minute-level** accelerometry data most popular for **digital biomarker development**

➢ Well-defined, **open-source, reproducible** methods to summarize minute-level accelerometry data, **non-overlapping** time window

    ➢ Monitor Independent Movement Summary (MIMS)
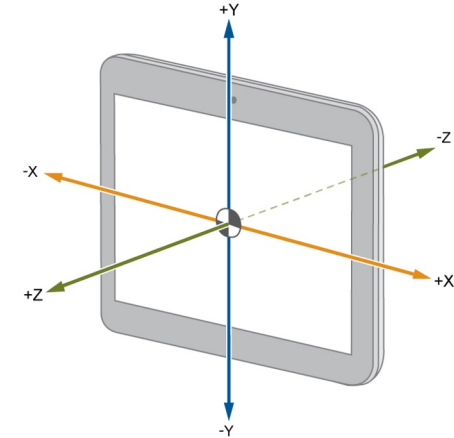
Weekly physical activity patterns suggest active and sedentary time within a day, and differences between weekdays and **weekends**
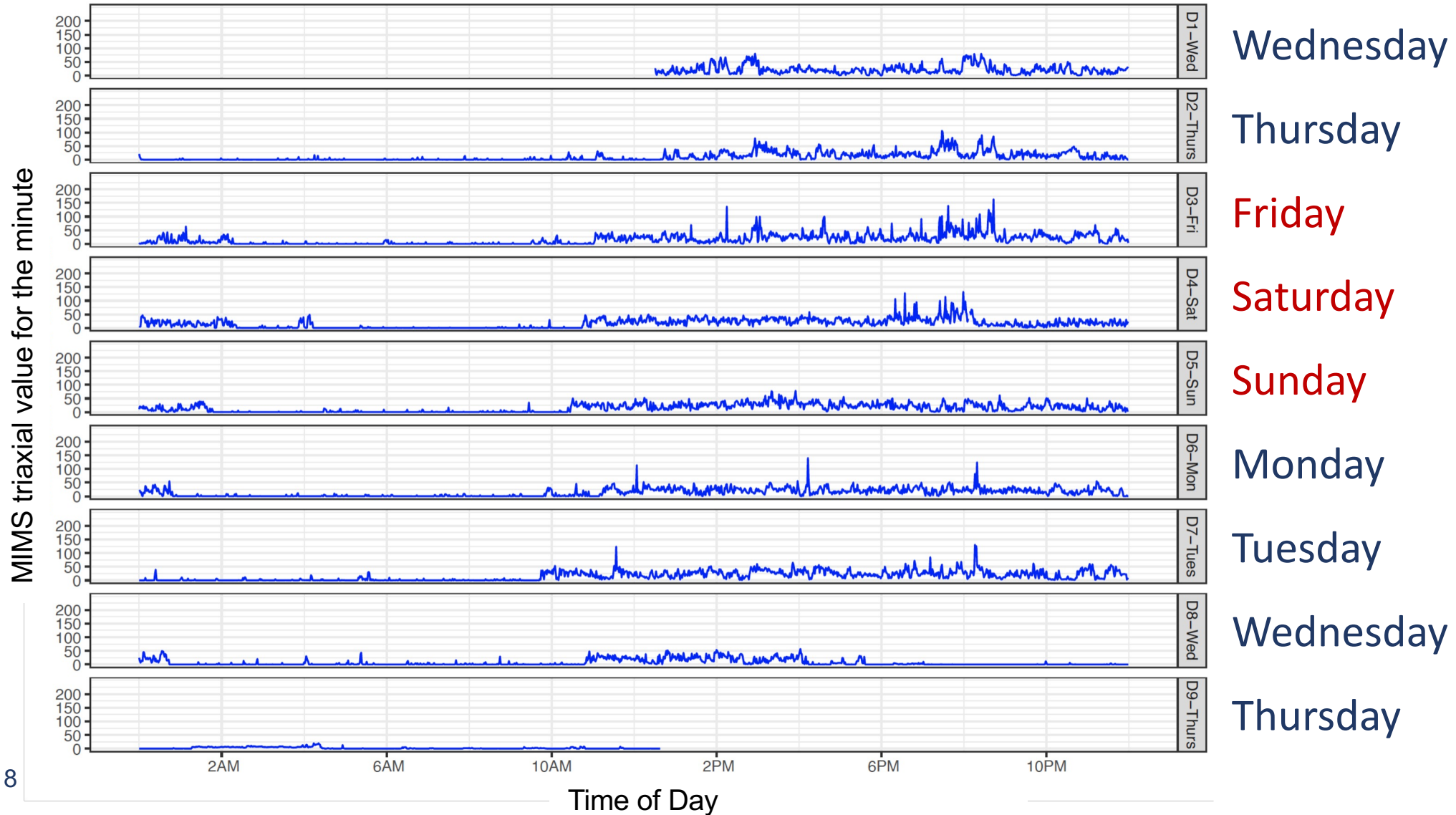
Arctools R package computes physical activity summaries

- ➢ *Arctools R* package processes accelerometry **minute-level data** to quantify physical activity characteristics:
    - ➢ Control data quality (1)
    - ➢ Impute missing data (2)

- ➢ (1) *Arctools* algorithm to **classify wear/non-wear**:
    - ➢ 90 or more minutes of consecutive 0's: non-wear
    - ➢ 144 or more minutes of consecutives 0's: invalid day
    - ➢ 2 or fewer valid days: non-representative subject

Arctools package created by *Dr. Jacek Urbanek*
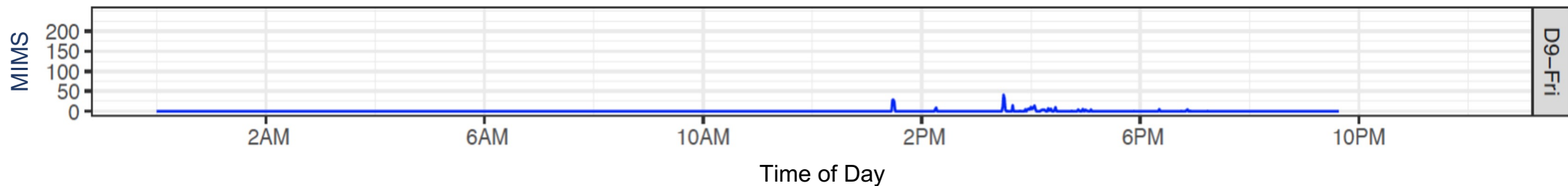
Arctools package imputes missing data with **median** physical activity
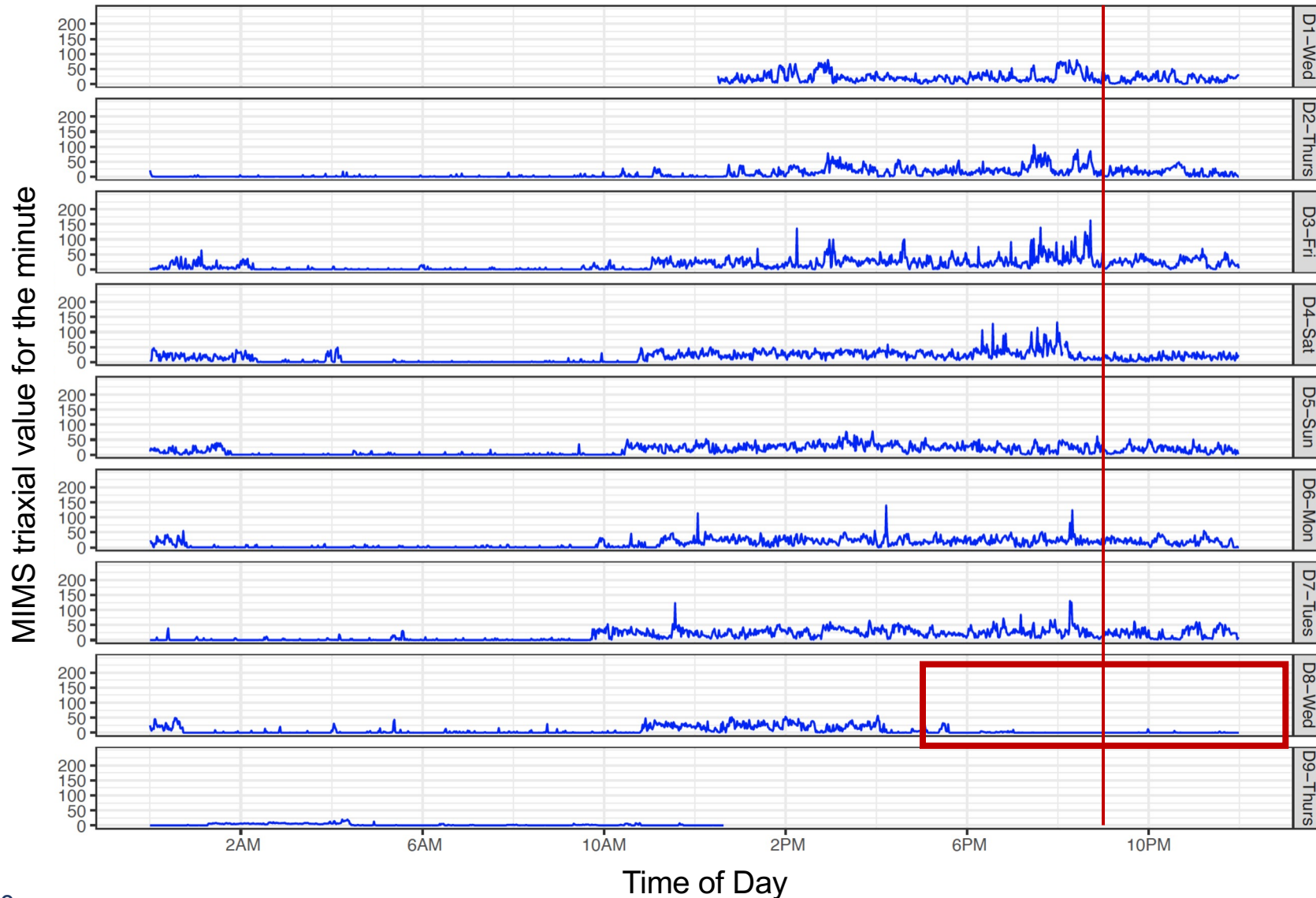
> ➤ (2) *Arctools* imputes **missing data** with median physical activity value from valid days within one person

Missing data

Arctools quantifies physical activity characteristics

➢ ***Arctools* output:**

    ➢ Minute-level metric of physical activity (PA) volume (AC, MIMS, ENMO, etc.)

    ➢ **Active minute**: a minute with AC equal or above a fixed universal threshold

    ➢ **Sedentary minute**: a minute with AC below a fixed universal threshold

    ➢ N valid days: number of days with ≤10% of non-wear time

    ➢ N days: number of unique day dates in the data collection

    ➢ Wear time on valid days: average number of wear time minutes across valid days

MIMS threshold: 10.558

Time of Day
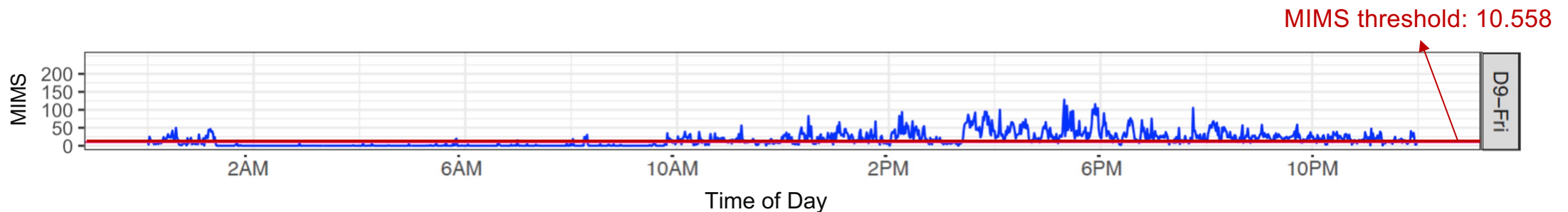
PA metrics from *Arctools* package contain much of the same information with minor variations
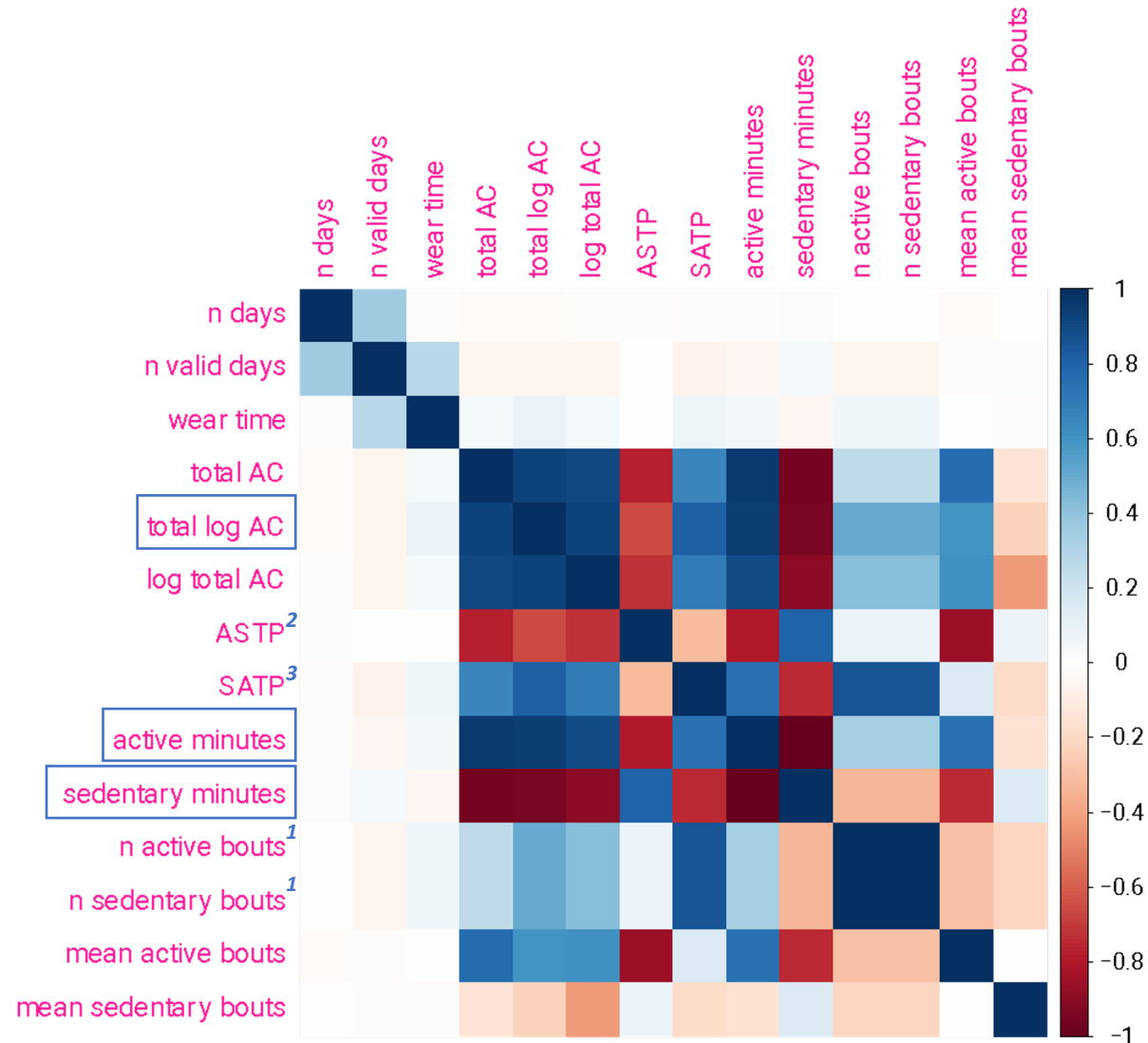
- ➢ (1) Active/sedentary bout: a sequence of one or more consecutive active/sedentary minutes
- ➢ (2) ASTP: active to sedentary transition probability
- ➢ (3) SATP: sedentary to active transition probability
- ➢ **Our focus: Total log Activity Count**

PA: physical activity

National Health and Nutrition Examination Survey (NHANES)

➢ National CDC survey conducted every 2 years

➢ Designed to assess health and nutritional status of adults and children in the US

➢ Survey takes form of interviews and physical examinations

➢ Oversampled traditionally underrepresented groups

➢ Includes demographic, socioeconomic, medical history and examination data

➢ NHANES is free and open access!
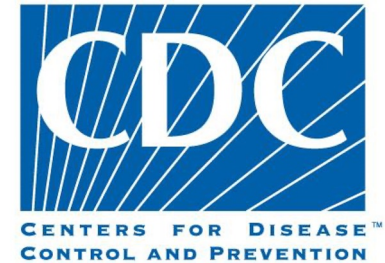
Study sample includes two NHANES cohorts 2011-2012 and 2013-2014

➢ Wear Physical Activity Monitor (PAM)

➢ Device manufactured by ActiGraph LLC

➢ Sampling frequency is at 80 Hz

➢ PA amplitude recorded on triaxial scale (MIMS scale)

PA: physical activity
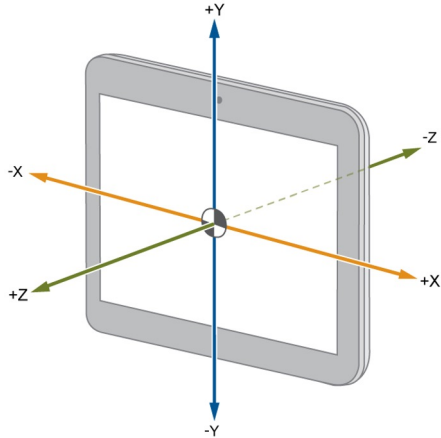
NHANES is a well-characterized data source

- ➢ **Demographics information**
  - ➢ Body Mass Index (kg/m^2)
  - ➢ Biological Sex
  - ➢ Age
  - ➢ Race
  - ➢ Highest education level
  - ➢ Marital status for people 20+
  - ➢ Pregnancy status
  - ➢ Annual household income groups
  - ➢ Number of people in the participant's household

- ➢ **Medical history**
  - ➢ Diabetes
  - ➢ Asthma
  - ➢ Congestive heart failure
  - ➢ Cancer
  - ➢ Stroke
  - ➢ Liver condition

- ➢ **Physical Examination data**
  - ➢ Spirometry
  - ➢ Audiology

- ➢ **And many more!**

NHANES Flow Chart

Overall sample (n=18,378)

With ActiGraphy (n=14,693)

No ActiGraphy (n=3,685)

Exclusion criteria (chronologically):
- Age <= 18 (n = 4,923)
- BMI >= 60 (n = 26)
- Valid days <= 2 (n = 841)

Study sample (n=8,903)
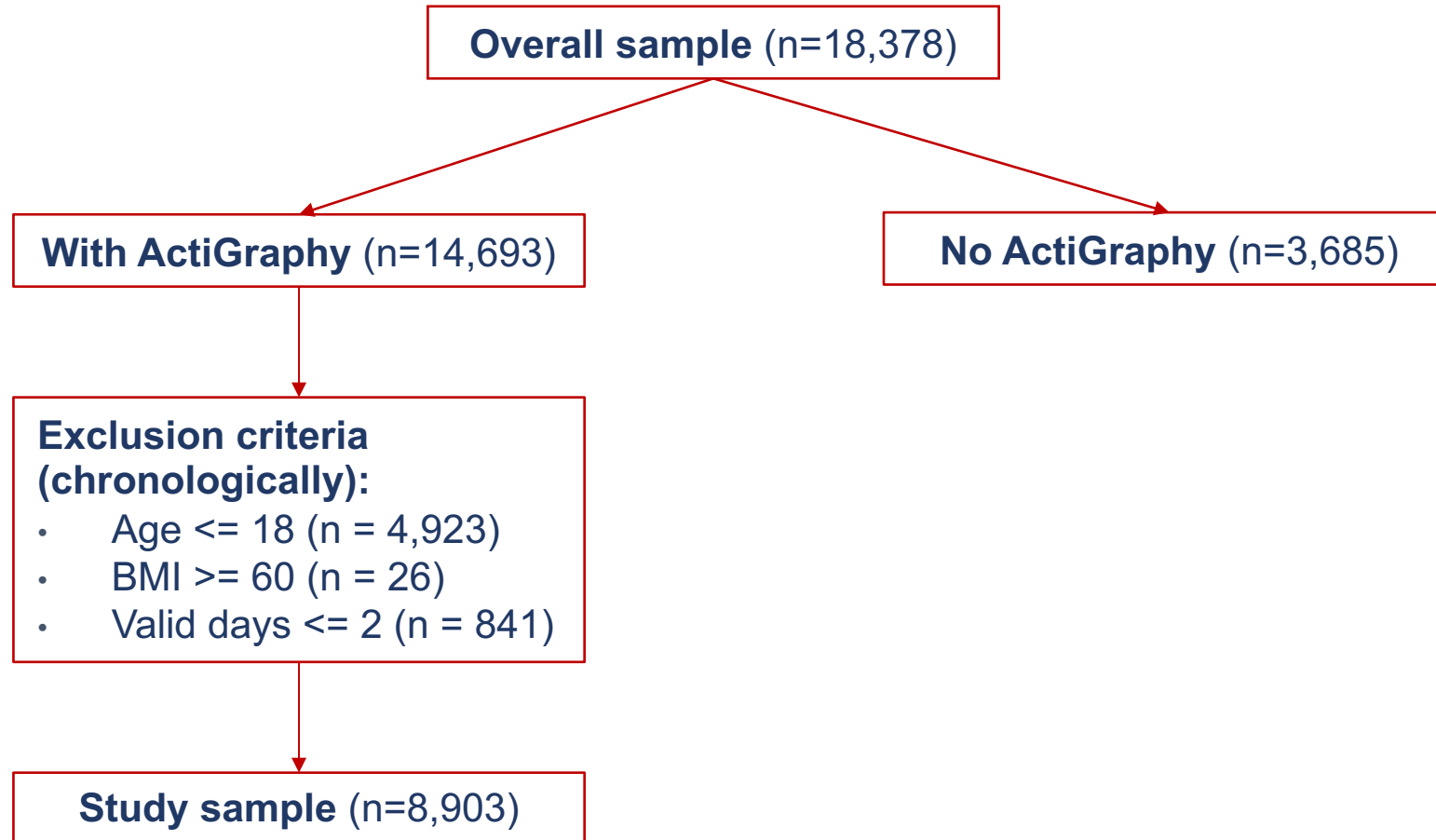
The two cohorts have similar physical activity estimates

	NHANES adults with Actigraphy	Cohort	
Characteristic	**Overall**, N = 8,903	**2011-2012**, N = 4,517	**2013-2014**, N = 4,386
Total log AC per valid day			
Mean (SD)	2,231 (438)	2,185 (479)	2,279 (385)
Median (IQR)	2,273 (2,003, 2,519)	2,244 (1,968, 2,491)	2,303 (2,037, 2,545)
Range	29, 3,974	29, 3,974	40, 3,942
Unknown	0 (0%)	0 (0%)	0 (0%)
Average number of active minutes per valid day			
Mean (SD)	482 (148)	471 (153)	492 (141)
Median (IQR)	486 (385, 586)	481 (377, 576)	492 (395, 596)
Range	0, 1,177	0, 1,177	0, 990
Unknown	0 (0%)	0 (0%)	0 (0%)
Average number of sedentary minutes per valid day			
Mean (SD)	957 (147)	968 (153)	947 (141)
Median (IQR)	953 (853, 1,054)	958 (863, 1,062)	947 (844, 1,044)
Range	263, 1,440	263, 1,440	450, 1,440
Unknown	0 (0%)	0 (0%)	0 (0%)

Two cohorts have comparable PA characteristics

PA: physical activity

Median age was 49 years and approximately 37% of the sample was obese

NHANES adults with Actigraphy

Characteristic	Overall, N = 8,903[1]	Cohort	
		2011-2012, N = 4,517[1]	2013-2014, N = 4,386[1]
age in years			
Mean (SD)	49 (18)	49 (18)	50 (18)
Median (IQR)	49 (34, 64)	49 (33, 63)	50 (35, 64)
Range	19, 80	19, 80	19, 80
Unknown	0 (0%)	0 (0%)	0 (0%)
BMI Categories			
severely underweight	16 (0.2%)	9 (0.2%)	7 (0.2%)
underweight	159 (1.8%)	88 (1.9%)	71 (1.6%)
normal weight	2,491 (28%)	1,300 (29%)	1,191 (27%)
overweight	2,806 (32%)	1,416 (31%)	1,390 (32%)
obesity class I	1,874 (21%)	930 (21%)	944 (22%)
obesity class II	840 (9.4%)	406 (9.0%)	434 (9.9%)
obesity class III	621 (7.0%)	310 (6.9%)	311 (7.1%)
Unknown	96 (1.1%)	58 (1.3%)	38 (0.9%)

[1] n (%)

The NHANES study oversampled black individuals

NHANES adults with Actigraphy

Characteristic	Overall, N = 8,903[1]	Cohort 2011-2012, N = 4,517[1]	2013-2014, N = 4,386[1]
biological sex			
female	4,634 (52%)	2,306 (51%)	2,328 (53%)
male	4,269 (48%)	2,211 (49%)	2,058 (47%)
race			
Asian	1,032 (12%)	585 (13%)	447 (10%)
White	3,599 (40%)	1,677 (37%)	1,922 (44%)
Black	2,124 (24%)	1,222 (27%)	902 (21%)
Hispanic	1,887 (21%)	905 (20%)	982 (22%)
Other Race	261 (2.9%)	128 (2.8%)	133 (3.0%)
Unknown	0 (0%)	0 (0%)	0 (0%)
annual household income groups			
$0 to $24,999	2,605 (29%)	1,417 (31%)	1,188 (27%)
$25,000 to $44,999	1,779 (20%)	906 (20%)	873 (20%)
$45,000 to $74,999	1,467 (16%)	685 (15%)	782 (18%)
$75,000 to $99,999	746 (8.4%)	374 (8.3%)	372 (8.5%)
$100,000 and over	1,509 (17%)	726 (16%)	783 (18%)
Unknown	797 (9.0%)	409 (9.1%)	388 (8.8%)

[1] n (%)

Demographics of two NHANES cohorts

NHANES adults with Actigraphy

Characteristic	Overall, N = 8,903[1]	Cohort 2011-2012, N = 4,517[1]	2013-2014, N = 4,386[1]
highest education level			
less than 9th grade	746 (8.4%)	409 (9.1%)	337 (7.7%)
9-11th grade	1,176 (13%)	599 (13%)	577 (13%)
high school graduate	1,902 (21%)	934 (21%)	968 (22%)
college or AA degree	2,636 (30%)	1,315 (29%)	1,321 (30%)
college graduate or above	2,206 (25%)	1,134 (25%)	1,072 (24%)
Unknown	237 (2.7%)	126 (2.8%)	111 (2.5%)
number of people in the participant's family			
1	1,248 (14%)	638 (14%)	610 (14%)
2	2,594 (29%)	1,383 (31%)	1,211 (28%)
3	1,615 (18%)	823 (18%)	792 (18%)
4	1,470 (17%)	752 (17%)	718 (16%)
5	1,976 (22%)	921 (20%)	1,055 (24%)
number of children aged 5 years or younger			
0	7,150 (80%)	3,658 (81%)	3,492 (80%)
1	1,234 (14%)	611 (14%)	623 (14%)
2	427 (4.8%)	213 (4.7%)	214 (4.9%)
3	92 (1.0%)	35 (0.8%)	57 (1.3%)

[1] n (%)

The ratio of family income to poverty was ~2:1

NHANES adults with Actigraphy

Characteristic	Overall, N = 8,903[1]	Cohort 2011-2012, N = 4,517[1]	2013-2014, N = 4,386[1]
six-month time period			
May 1 through October 31	4,561 (51%)	2,354 (52%)	2,207 (50%)
November 1 through April 30	4,342 (49%)	2,163 (48%)	2,179 (50%)
pregnancy status, for women 20-44 yrs old			
no	1,689 (19%)	849 (19%)	840 (19%)
yes	85 (1.0%)	40 (0.9%)	45 (1.0%)
Unknown	7,129 (80%)	3,628 (80%)	3,501 (80%)
ratio of family income to poverty			
Mean (SD)	2.44 (1.65)	2.40 (1.67)	2.47 (1.64)
Median (IQR)	1.97 (1.02, 3.97)	1.88 (0.98, 3.97)	2.06 (1.04, 3.93)
Range	0.00, 5.00	0.00, 5.00	0.00, 5.00
Unknown	659 (7.4%)	345 (7.6%)	314 (7.2%)

[1] n (%)

21

Self-reported diabetes, asthma, and cancer were common health conditions in the cohort

NHANES adults with Actigraphy

Characteristic	Overall, N = 8,903[1]	Cohort 2011-2012, N = 4,517[1]	2013-2014, N = 4,386[1]
health condition			
excellent	783 (8.8%)	414 (9.2%)	369 (8.4%)
very good	2,225 (25%)	1,138 (25%)	1,087 (25%)
good	3,344 (38%)	1,653 (37%)	1,691 (39%)
fair	303 (3.4%)	151 (3.3%)	152 (3.5%)
poor	1,648 (19%)	795 (18%)	853 (19%)
Unknown	600 (6.7%)	366 (8.1%)	234 (5.3%)
diabetes			
no	7,496 (84%)	3,832 (85%)	3,664 (84%)
borderline	238 (2.7%)	100 (2.2%)	138 (3.1%)
yes	1,164 (13%)	583 (13%)	581 (13%)
Unknown	5 (<0.1%)	2 (<0.1%)	3 (<0.1%)
asthma			
no	7,553 (85%)	3,833 (85%)	3,720 (85%)
yes	1,342 (15%)	680 (15%)	662 (15%)
Unknown	8 (<0.1%)	4 (<0.1%)	4 (<0.1%)

NHANES adults with Actigraphy

| Characteristic	Overall, N = 8,903[1]	Cohort 2011-2012, N = 4,517[1]	2013-2014, N = 4,386[1]
congestive heart failure			
no	8,356 (94%)	4,222 (93%)	4,134 (94%)
yes	301 (3.4%)	160 (3.5%)	141 (3.2%)
Unknown	246 (2.8%)	135 (3.0%)	111 (2.5%)
cancer			
no	7,830 (88%)	4,002 (89%)	3,828 (87%)
yes	842 (9.5%)	390 (8.6%)	452 (10%)
Unknown	231 (2.6%)	125 (2.8%)	106 (2.4%)
stroke			
no	8,321 (93%)	4,196 (93%)	4,125 (94%)
yes	346 (3.9%)	196 (4.3%)	150 (3.4%)
Unknown	236 (2.7%)	125 (2.8%)	111 (2.5%)
liver condition			
no	8,309 (93%)	4,215 (93%)	4,094 (93%)
yes	355 (4.0%)	176 (3.9%)	179 (4.1%)
Unknown	239 (2.7%)	126 (2.8%)	113 (2.6%)

Total log activity counts

Metric of PA volume and it is normally distributed
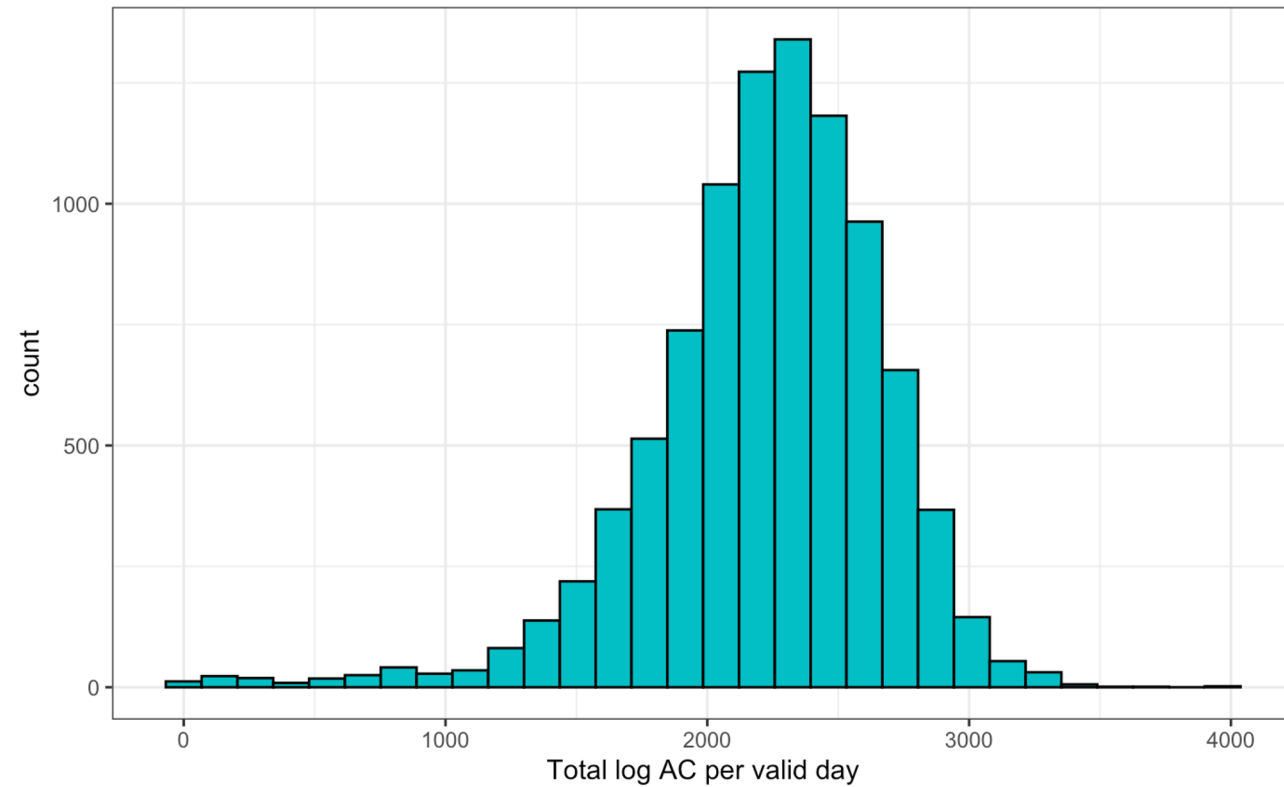
| Characteristic | Overall, N = 8,901 | Cohort | |
		2011-2012, N = 4,516	2013-2014, N = 4,385
Total log AC per valid day			
Mean (SD)	2,231 (437)	2,184 (479)	2,278 (384)
Median (IQR)	2,273 (2,003, 2,519)	2,244 (1,968, 2,491)	2,303 (2,037, 2,545)
Range	29, 3,562	29, 3,562	40, 3,343

2 subjects with Activity Count above 3800 units were removed for graphics

Scatterplot shows up and downward trend in age
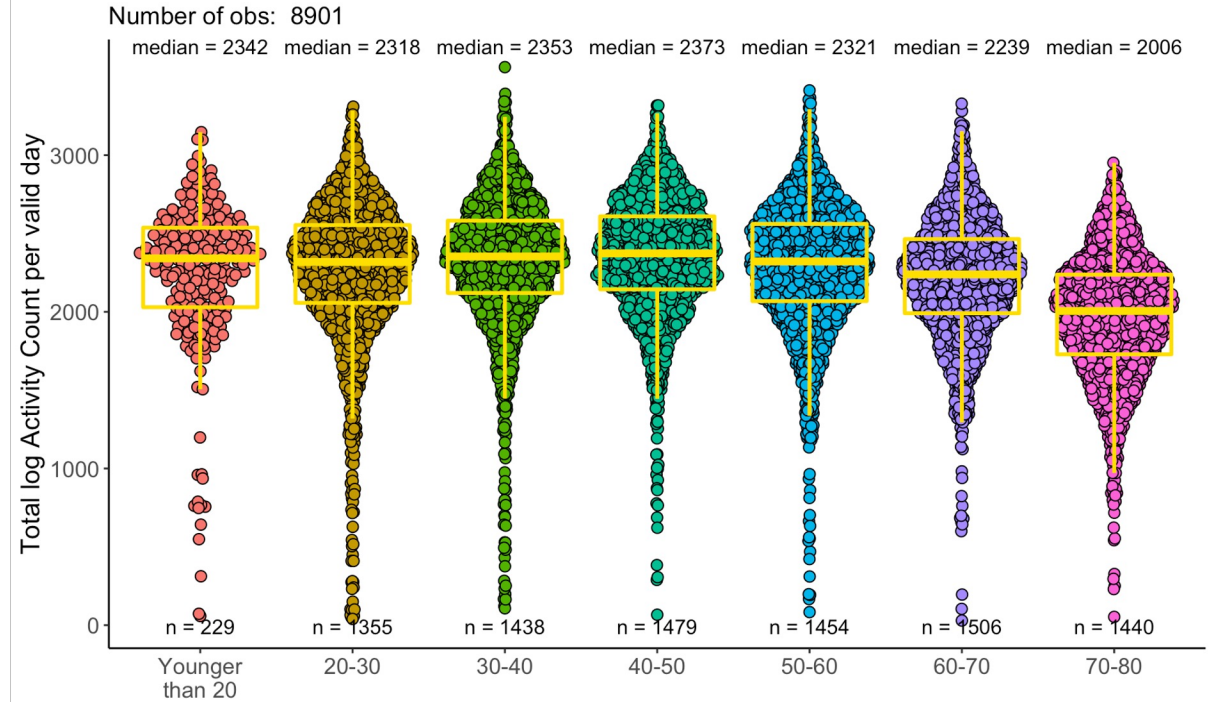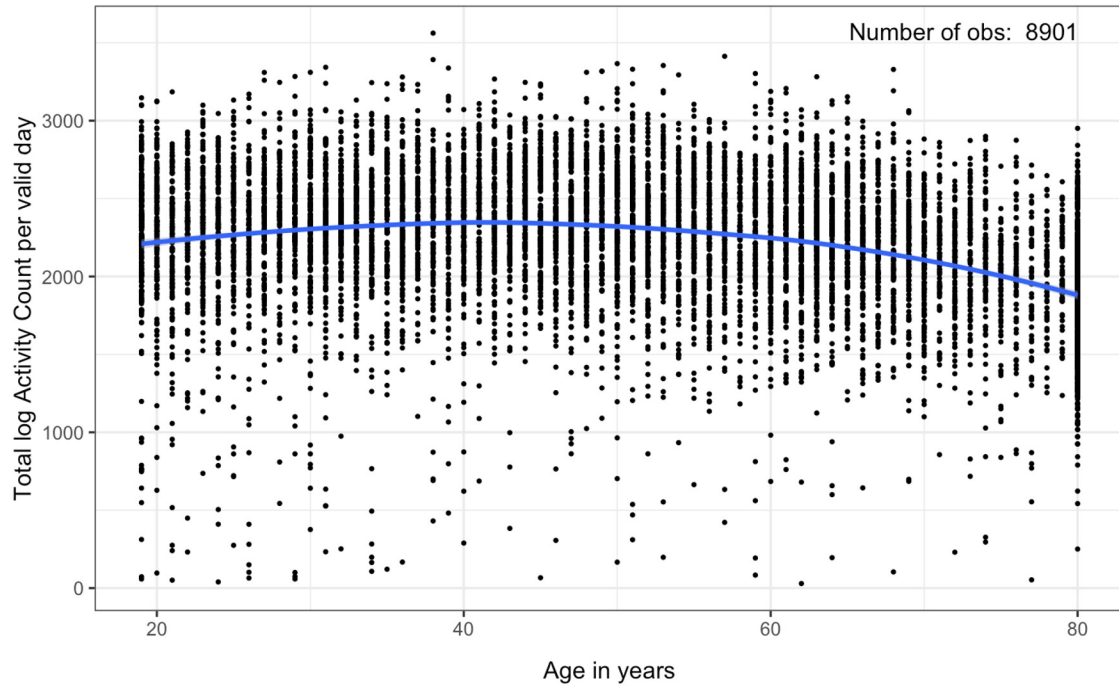
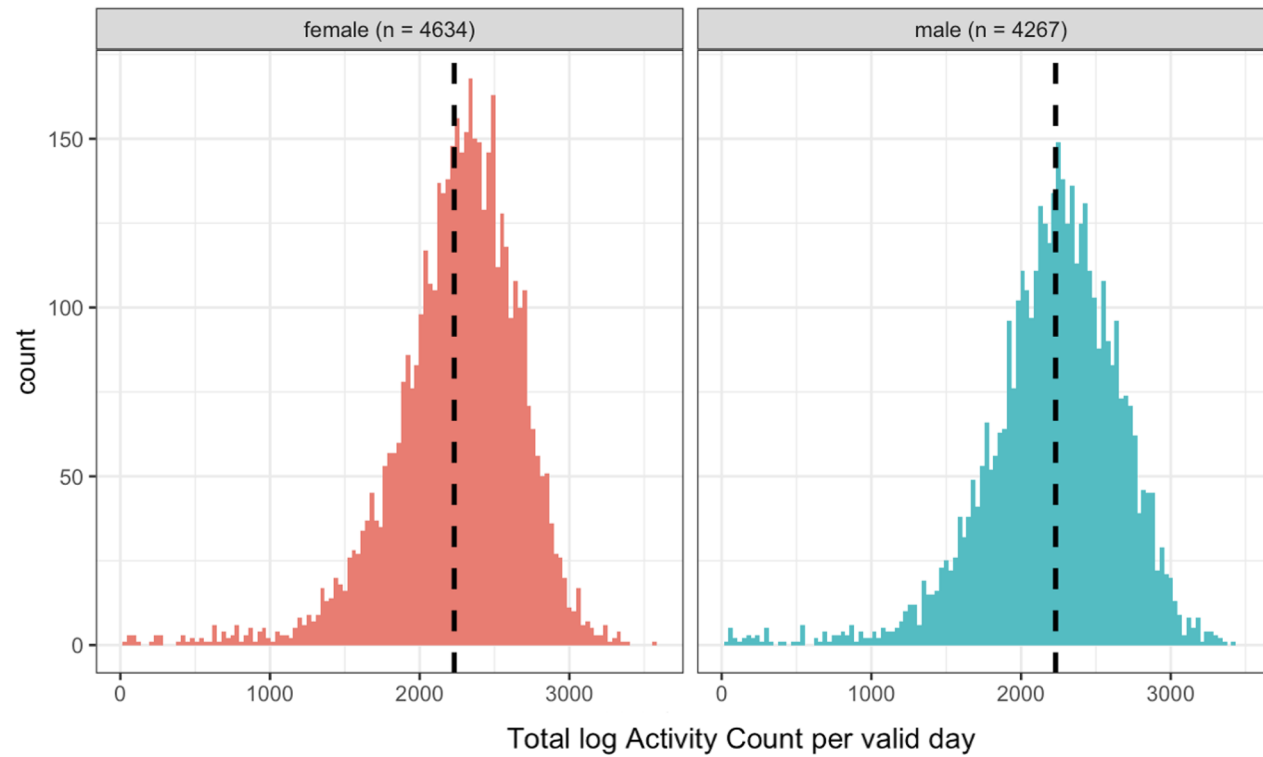| Characteristic	Younger than 20, N = 229[1]	20-30, N = 1,355[1]	30-40, N = 1,438[1]	40-50, N = 1,479[1]	50-60, N = 1,454[1]	60-70, N = 1,506[1]
Total log Activity Count per day						
Mean (SD)	2,234 (510)	2,254 (470)	2,316 (432)	2,349 (386)	2,283 (418)	2,210 (396)
Median (IQR)	2,342 (2,029, 2,537)	2,318 (2,057, 2,554)	2,353 (2,120, 2,582)	2,373 (2,144, 2,610)	2,321 (2,068, 2,561)	2,239 (1,991, 2,464)
Range	58, 3,147	40, 3,310	108, 3,562	67, 3,318	83, 3,413	29, 3,329

2 subjects with Activity Count above 3800 units were removed for graphics

Scatterplot shows U shape of physical activity and BMI

Characteristic	severely underweight, N = 16	underweight, N = 159	normal weight, N = 2,491	overweight, N = 2,805	obesity class I, N = 1,874	obesity class II, N = 840	obesity class III, N = 620
Total log AC per day							
Mean (SD)	1,873 (398)	2,118 (507)	2,256 (447)	2,249 (432)	2,235 (414)	2,210 (425)	2,150 (430)
Median (IQR)	2,018 (1,768, 2,116)	2,160 (1,914, 2,404)	2,304 (2,050, 2,539)	2,287 (2,014, 2,540)	2,266 (2,008, 2,513)	2,240 (1,969, 2,506)	2,195 (1,921, 2,423)
Range	857, 2,329	65, 3,207	58, 3,562	40, 3,318	29, 3,392	83, 3,366	53, 3,201

- Severely underweight - BMI < 16.5
- Underweight - BMI under 18.5
- Normal weight - BMI 18.5 to 24.9
- Overweight – BMI 25 to 29.9
- Obesity class I – BMI 30 to 34.9
- Obesity class II – BMI 35 to 39.9
- Obesity class III – BMI > 40

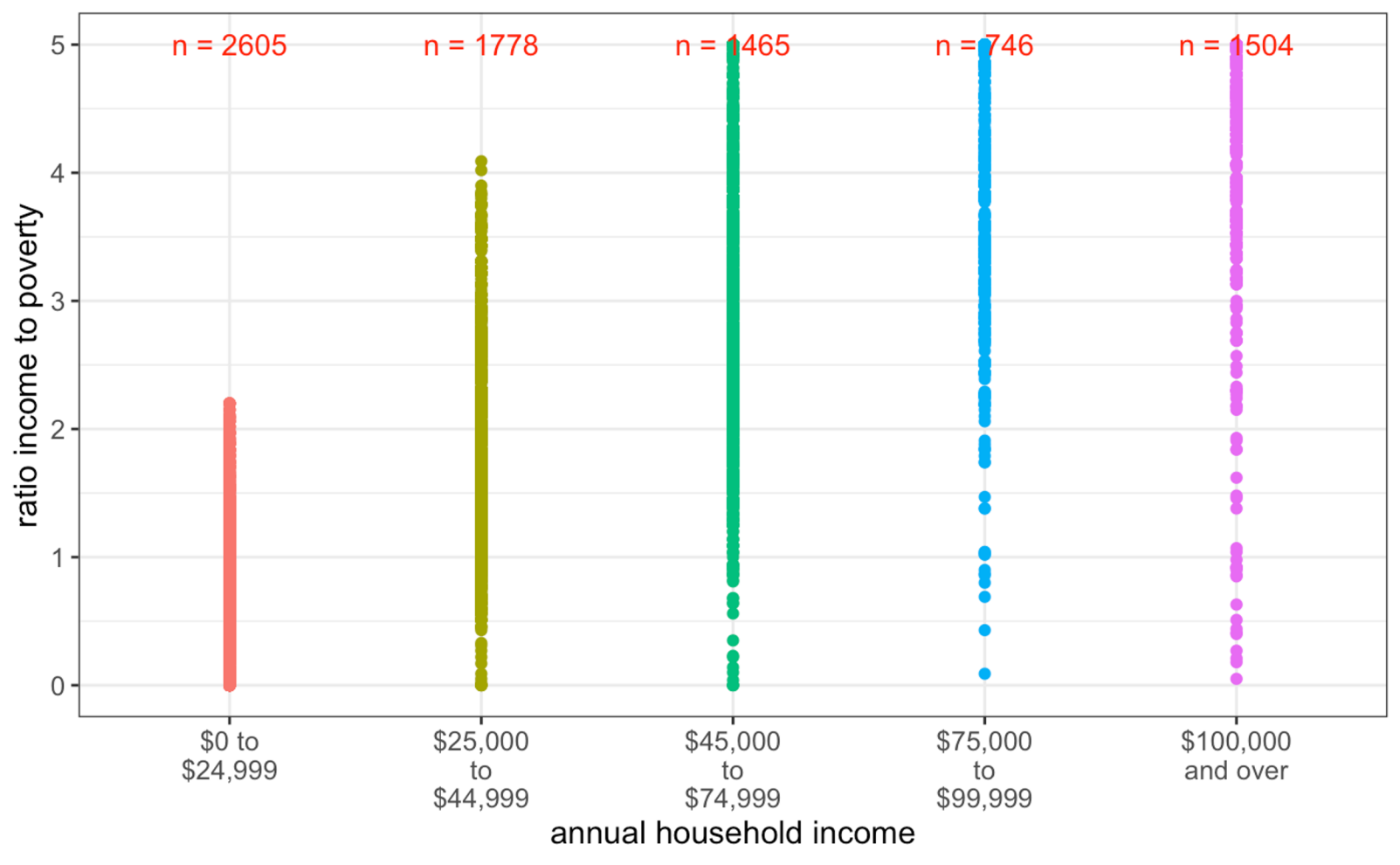*2 subjects with Activity Count above 3800 units were removed for graphics*

Men and women have similar physical activity distributions

Characteristic	female, N = 4,634	male, N = 4,267
Total log Activity Count per day		
Mean (SD)	2,248 (425)	2,212 (449)
Median (IQR)	2,291 (2,029, 2,529)	2,250 (1,979, 2,507)
Range	29, 3,562	40, 3,413

2 subjects with Activity Count above 3800 units were removed for graphics

Ratio of income to poverty strongly associated with income groups

Spearman Correlation = 0.88

Among ethnic groups, Hispanics have a higher distribution of physical activity

Number of obs: 8901
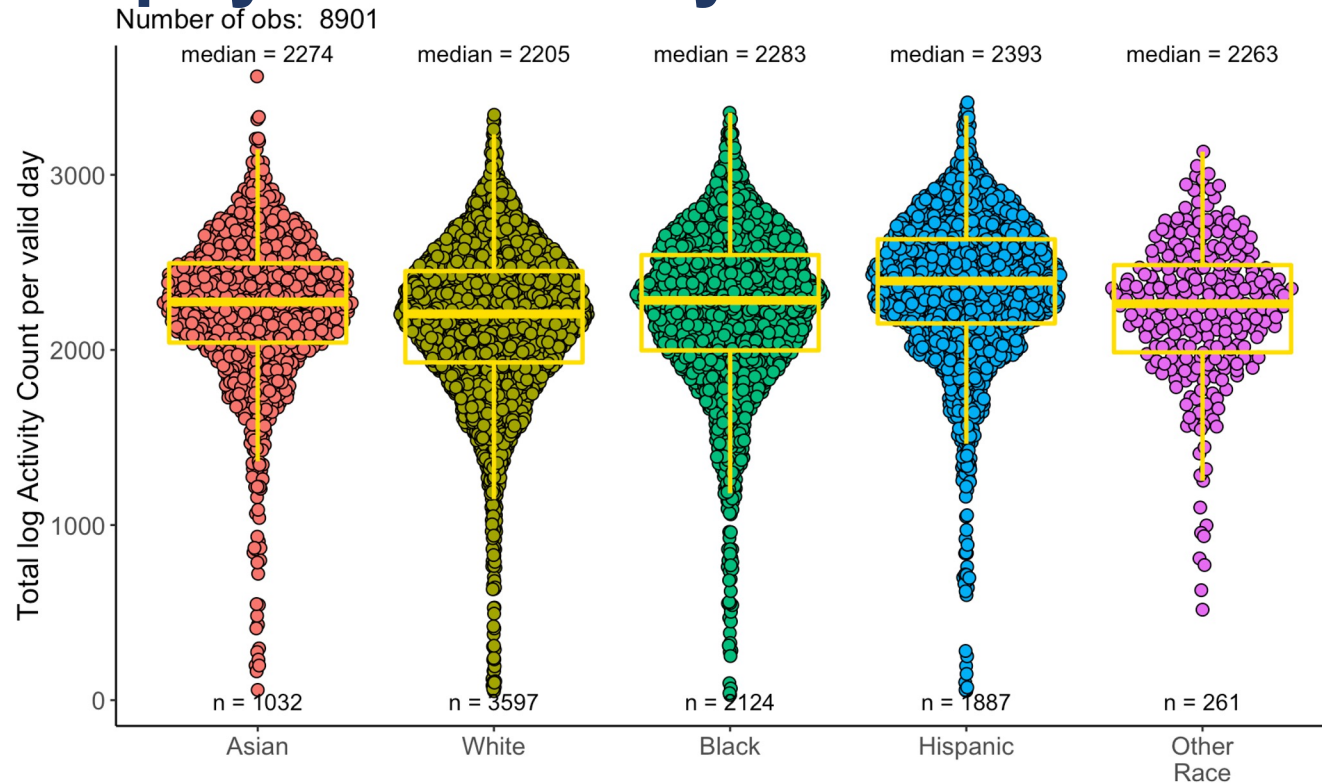
Characteristic	**Asian**, N = 1,032	**White**, N = 3,597	**Black**, N = 2,124	**Hispanic**, N = 1,887	**Other Race**, N = 261
Total log Activity Count per day					
Mean (SD)	2,231 (437)	2,166 (423)	2,232 (451)	2,353 (422)	2,218 (429)
Median (IQR)	2,274 (2,042, 2,495)	2,205 (1,929, 2,450)	2,283 (1,997, 2,542)	2,393 (2,151, 2,632)	2,263 (1,986, 2,484)
Range	59, 3,562	51, 3,343	29, 3,354	53, 3,413	517, 3,132

2 subjects with Activity Count above 3800 units were removed for graphics

No pattern in physical activity across educational levels
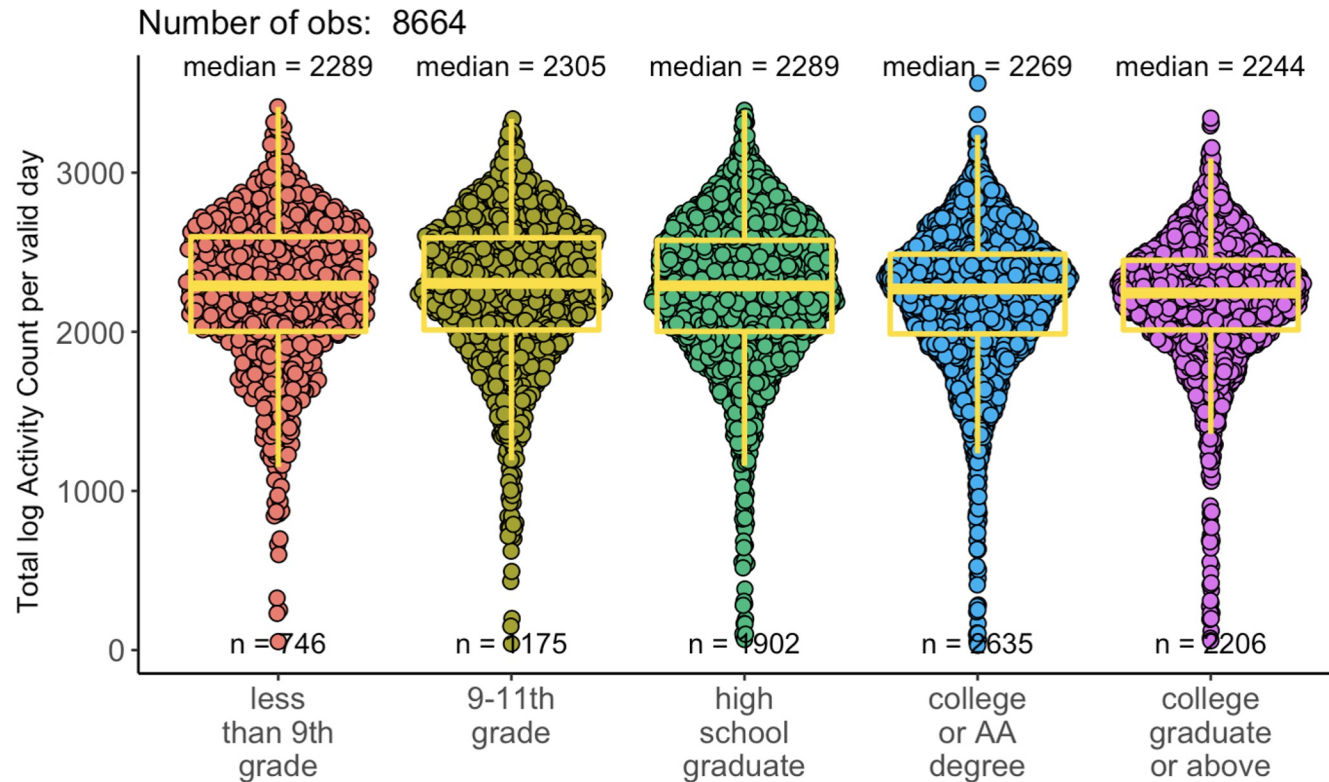
Characteristic	less than 9th grade, N = 746	9-11th grade, N = 1,175	high school graduate, N = 1,902	college or AA degree, N = 2,635	college graduate or above, N = 2,206
Total log Activity Count per day					
Mean (SD)	2,246 (485)	2,263 (464)	2,256 (461)	2,216 (421)	2,205 (388)
Median (IQR)	2,289 (2,002, 2,598)	2,305 (2,012, 2,591)	2,289 (2,002, 2,575)	2,269 (1,987, 2,488)	2,244 (2,012, 2,449)
Range	53, 3,413	40, 3,338	67, 3,392	29, 3,562	59, 3,343

29 *2 subjects with Activity Count above 3800 units were removed for graphics*

Widowed people move the least, followed by divorced individuals
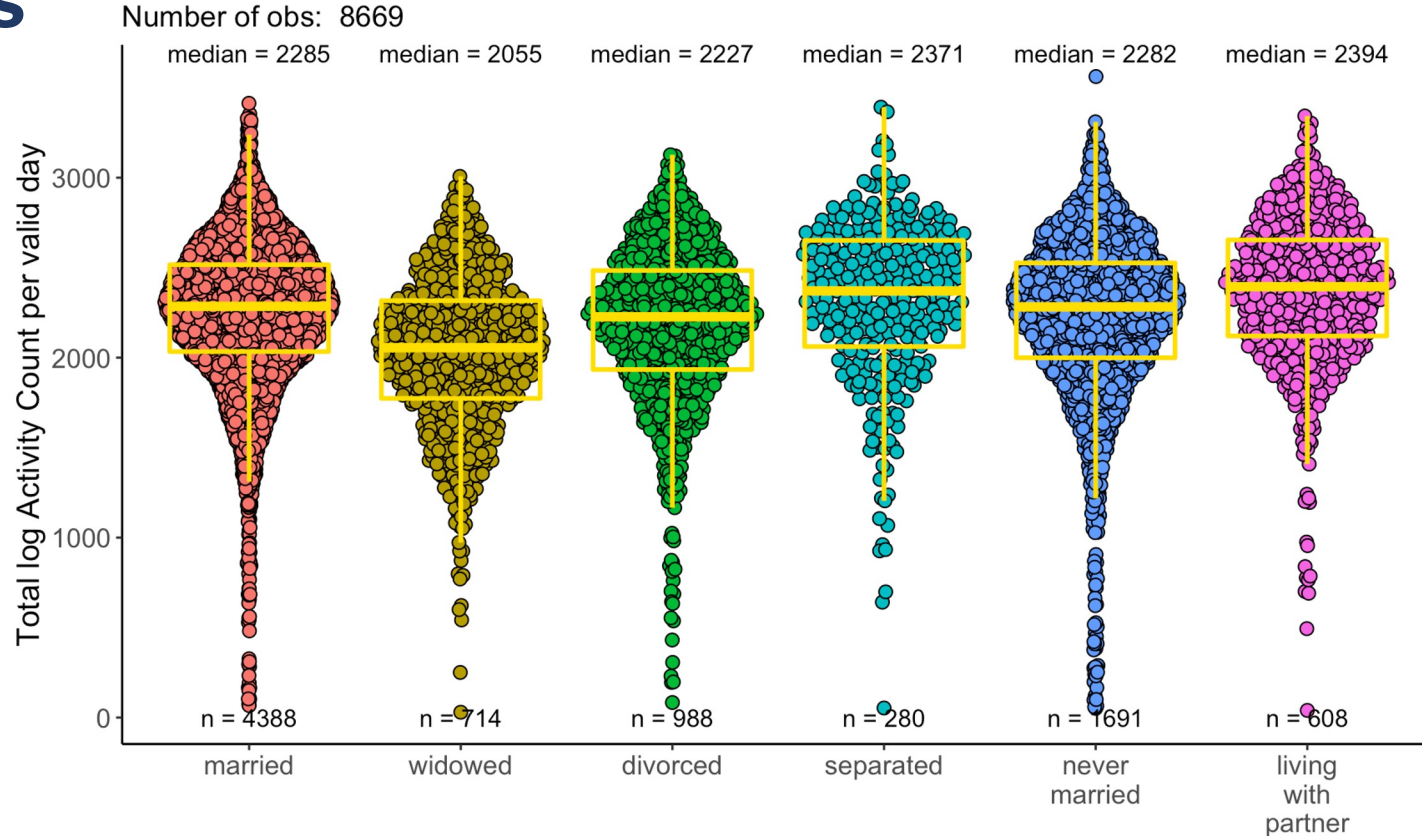
Number of obs: 8669

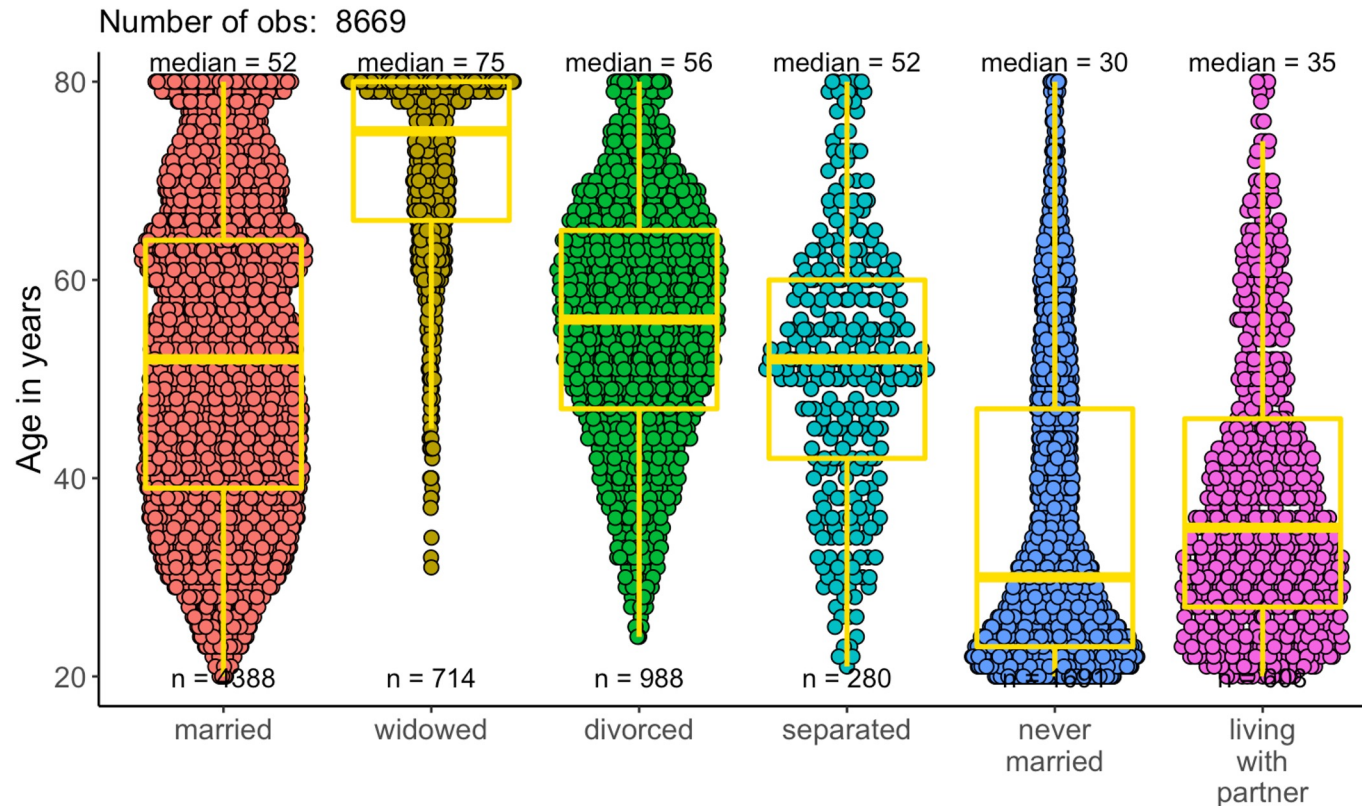| Characteristic	married, N = 4,388	widowed, N = 714	divorced, N = 988	separated, N = 280	never married, N = 1,691	living with partner, N = 608
Total log Activity Count per day						
Mean (SD)	2,253 (403)	2,026 (431)	2,184 (447)	2,303 (482)	2,227 (472)	2,363 (426)
Median (IQR)	2,285 (2,034, 2,517)	2,055 (1,774, 2,317)	2,227 (1,934, 2,484)	2,371 (2,062, 2,652)	2,282 (1,999, 2,526)	2,394 (2,120, 2,654)
Range	67, 3,413	29, 3,007	83, 3,127	53, 3,392	51, 3,562	40, 3,343

2 subjects with Activity Count above 3800 units were removed for graphics

Age distribution across marital groups


Number of obs: 8669

Characteristic	married, N = 4,388	widowed, N = 714	divorced, N = 988	separated, N = 280	never married, N = 1,691	living with partner, N = 608
age in years						
Mean (SD)	52 (16)	72 (10)	56 (12)	51 (14)	36 (15)	38 (14)
Median (IQR)	52 (39, 64)	75 (66, 80)	56 (47, 65)	52 (42, 60)	30 (23, 47)	35 (27, 46)
Range	20, 80	31, 80	24, 80	21, 80	20, 80	20, 80

2 subjects with Activity Count above 3800 units were removed for graphics

No physical activity differences by Asthmatic status
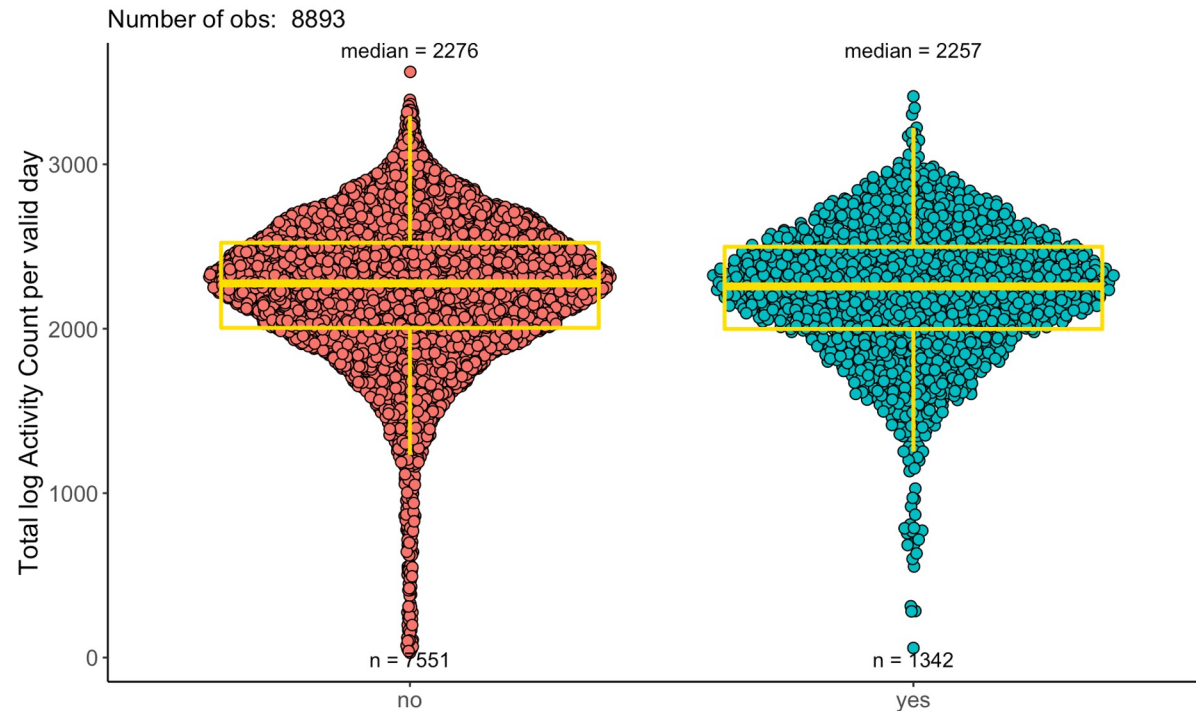
Characteristic	**no**, N = 7,551	**yes**, N = 1,342
Total log Activity Count per day		
Mean (SD)	2,233 (439)	2,219 (425)
Median (IQR)	2,276 (2,005, 2,523)	2,257 (1,999, 2,498)
Range	29, 3,562	58, 3,413

2 subjects with Activity Count above 3800 units were removed for graphics

People with self-reported past Cancer move less
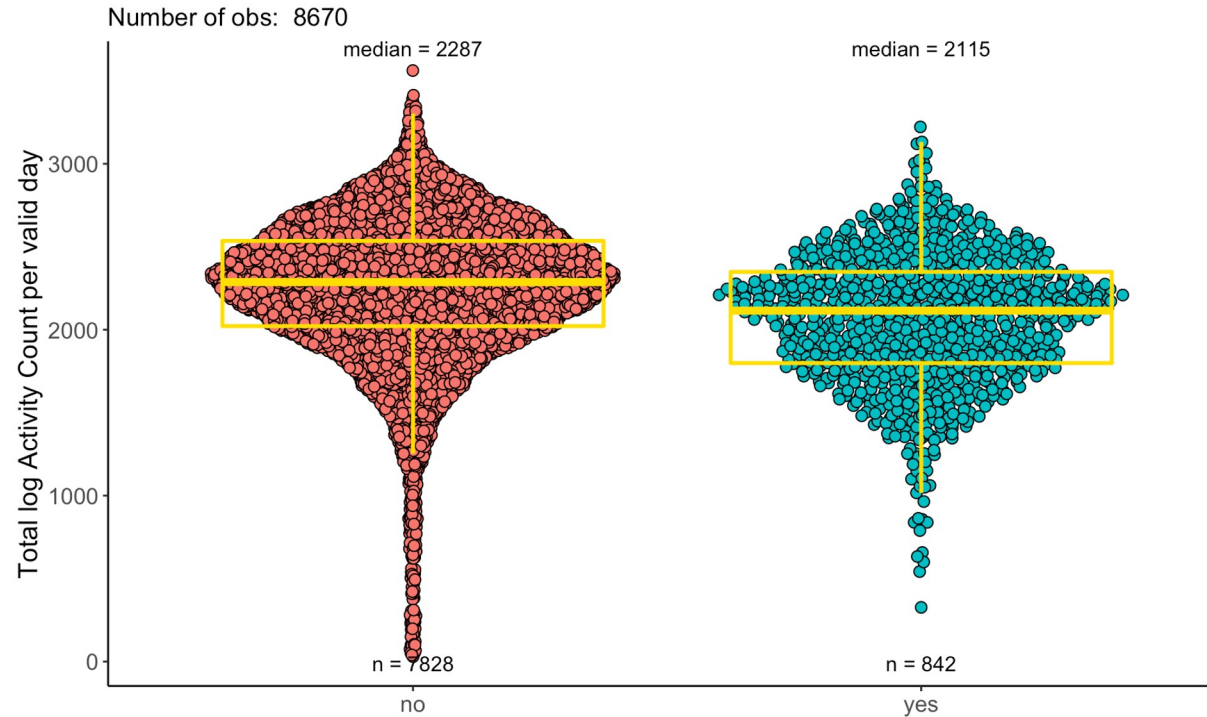
Characteristic	**no**, N = 7,828	**yes**, N = 842
Total log Activity Count per day		
Mean (SD)	2,248 (434)	2,070 (411)
Median (IQR)	2,287 (2,022, 2,535)	2,115 (1,799, 2,349)
Range	29, 3,562	327, 3,222

2 subjects with Activity Count above 3800 units were removed for graphics

Univariate analysis with activity counts

Significant p-values but small effect size

Significant p-values but the effect size is medium based on AC scale (AC median is 2273)

Univariate Linear Model

Characteristic	N	Beta	95% CI[1]	p-value
BMI Categories	8,807			<0.001
normal weight		—	—	
severely underweight		-384	-597, -170	
underweight		-138	-208, -69	
overweight		-6.5	-30, 17	
obesity class I		-21	-47, 4.8	
obesity class II		-46	-80, -12	
obesity class III		-103	-141, -65	
Body Mass Index (kg/m^2)	8,807	-3.0	-4.3, -1.6	<0.001
biological sex	8,903			<0.001
female		—	—	
male		-35	-53, -17	

[1] CI = Confidence Interval

Univariate Linear Model

Characteristic	N	Beta	95% CI[1]	p-value
race	8,903			<0.001
White		—	—	
Asian		64	34, 94	
Black		65	42, 88	
Hispanic		186	161, 210	
Other Race		51	-3.6, 105	
annual household income groups	8,106			<0.001
$0 to $24,999		—	—	
$25,000 to $44,999		50	24, 77	
$45,000 to $74,999		67	39, 95	
$75,000 to $99,999		81	46, 117	
$100,000 and over		58	30, 85	

[1] CI = Confidence Interval

34

Univariate analysis with activity counts

Significant p-values but small effect size

Ratio of family income to poverty is strongly associated with income groups (Spearman Correlation = 0.88)

Univariate Linear Model

Characteristic	N	Beta	95% CI[1]	p-value
race	8,903			<0.001
White		—	—	
Asian		64	34, 94	
Black		65	42, 88	
Hispanic		186	161, 210	
Other Race		51	-3.6, 105	
annual household income groups	8,106			<0.001
$0 to $24,999		—	—	
$25,000 to $44,999		50	24, 77	
$45,000 to $74,999		67	39, 95	
$75,000 to $99,999		81	46, 117	
$100,000 and over		58	30, 85	

[1] CI = Confidence Interval

Univariate Linear Model

Characteristic	N	Beta	95% CI[1]	p-value
age in 10 years	8,903	-50	-55, -45	<0.001
ratio of family income to poverty	8,244	0.25	-5.5, 6.0	>0.9
six-month time period	8,903			0.046
May 1 through October 31		—	—	
November 1 through April 30		-19	-37, -0.37	
education level among adults	8,666			<0.001
less than 9th grade		—	—	
9-11th grade		18	-22, 58	
high school graduate		9.7	-27, 47	
college or AA degree		-30	-65, 5.7	
college graduate or above		-42	-78, -5.4	

[1] CI = Confidence Interval

Univariate analysis with activity counts

Significant p-values but small effect size

Univariate Linear Model				
Characteristic	**N**	**Beta**	**95% CI[1]**	**p-value**
marital status for people > 20 yrs old	8,671			<0.001
married		—	—	
widowed		-227	-261, -193	
divorced		-69	-98, -39	
separated		51	-1.3, 103	
never married		-23	-47, 0.83	
living with partner		111	74, 147	
pregnancy status, for women 20-44 yrs old	1,774			0.026
no		—	—	
yes		-107	-201, -13	
number of people in the participant's family	8,903	64	57, 70	<0.001
number of children aged 5 years or younger	8,903	106	90, 121	<0.001
[1] CI = Confidence Interval				

Significant p-values, medium effect size based on AC scale (AC median is 2273)

Univariate analysis with activity counts

Significant p-values but small effect size

Significant p-values, medium effect size based on AC scale (AC median is 2273)

Univariate Linear Model

Characteristic	N	Beta	95% CI[1]	p-value
health condition	8,303			<0.001
poor		—	—	
excellent		89	52, 125	
very good		66	38, 93	
fair		-151	-204, -98	
good		48	22, 73	
asthma	8,895			0.3
no		—	—	
yes		-15	-40, 11	
cancer	8,672			<0.001
no		—	—	
yes		-178	-209, -147	

[1] CI = Confidence Interval

Univariate Linear Model

| Characteristic	N	Beta	95% CI[1]	p-value
diabetes	8,898			<0.001
no		—	—	
borderline		-39	-94, 17	
yes		-179	-206, -153	
congestive heart failure	8,657			<0.001
no		—	—	
yes		-267	-317, -217	
stroke	8,667			<0.001
no		—	—	
yes		-314	-360, -267	
liver condition	8,664			0.13
no		—	—	
yes		-36	-82, 11	

[1] CI = Confidence Interval

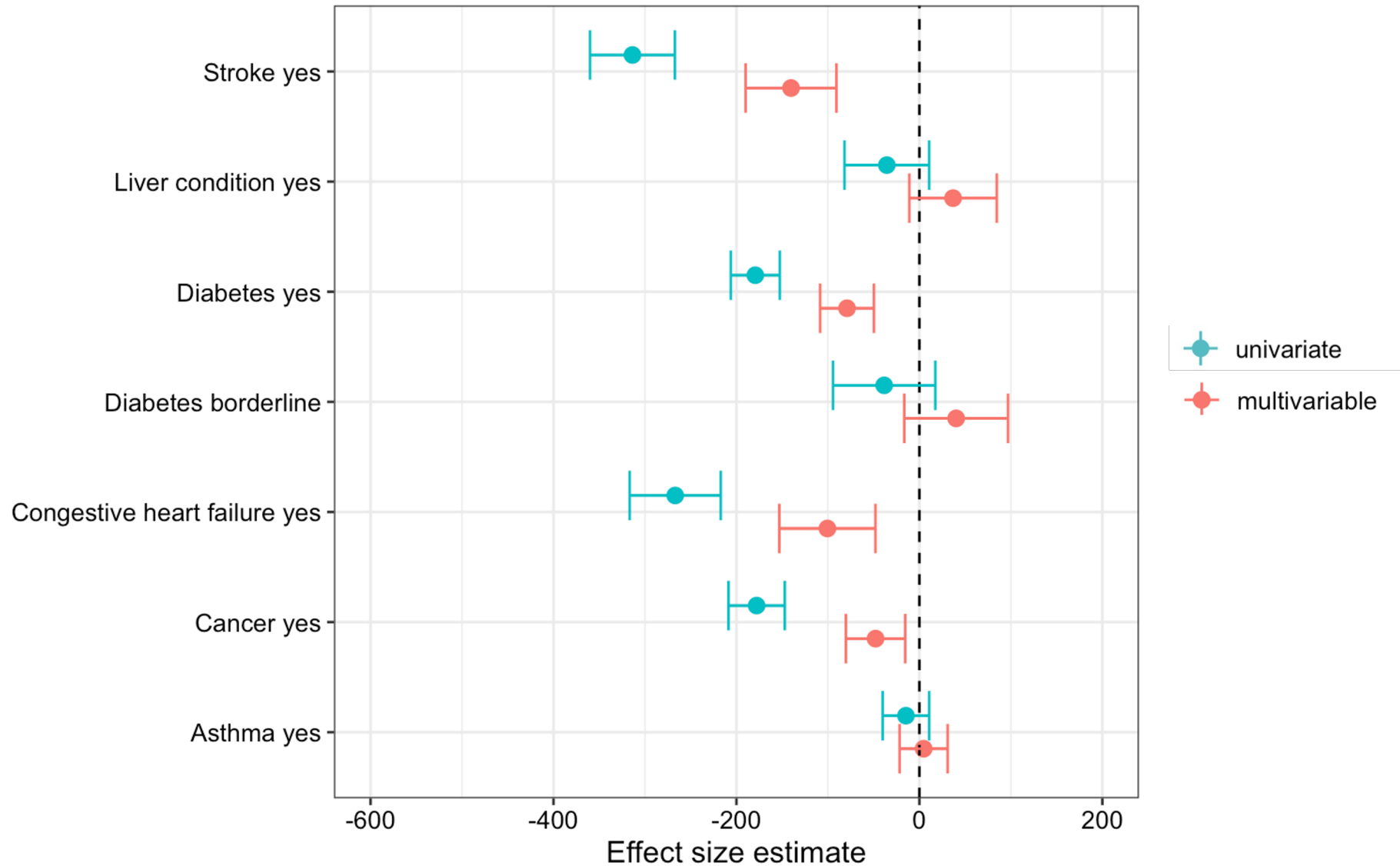# Forest plot of effect size for activity counts
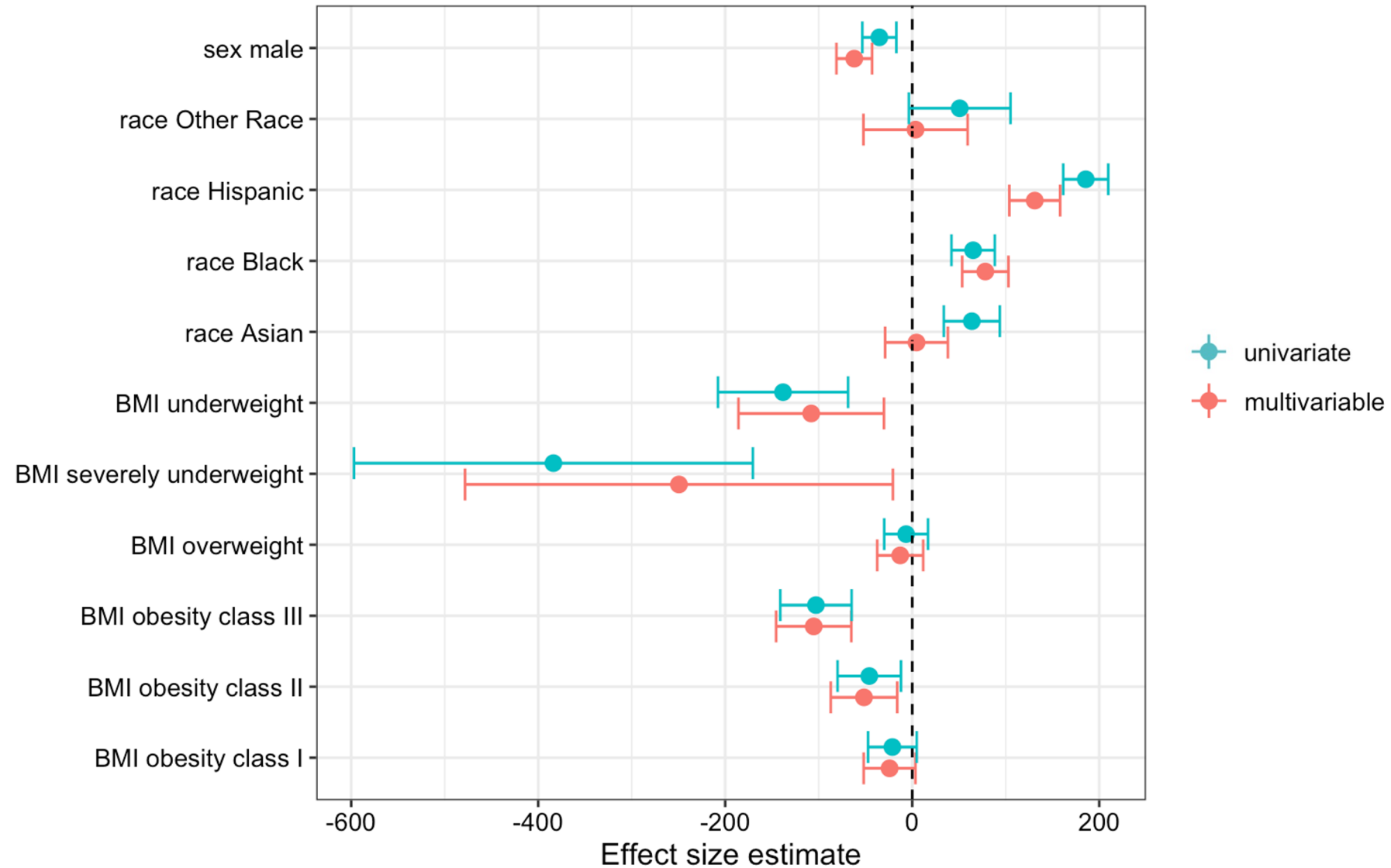
Forest plot of effect size for activity counts
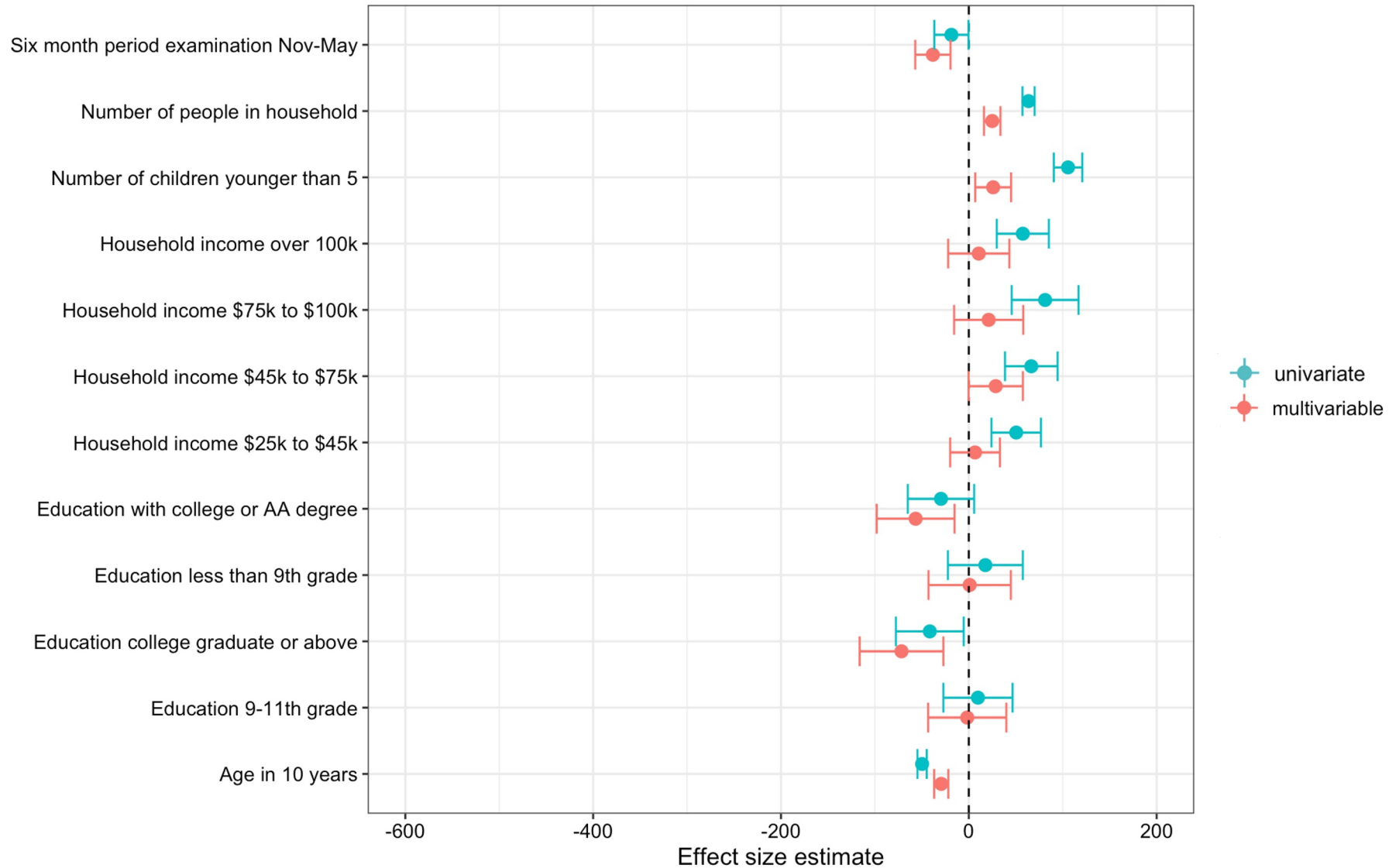
Forest plot of effect size for activity counts
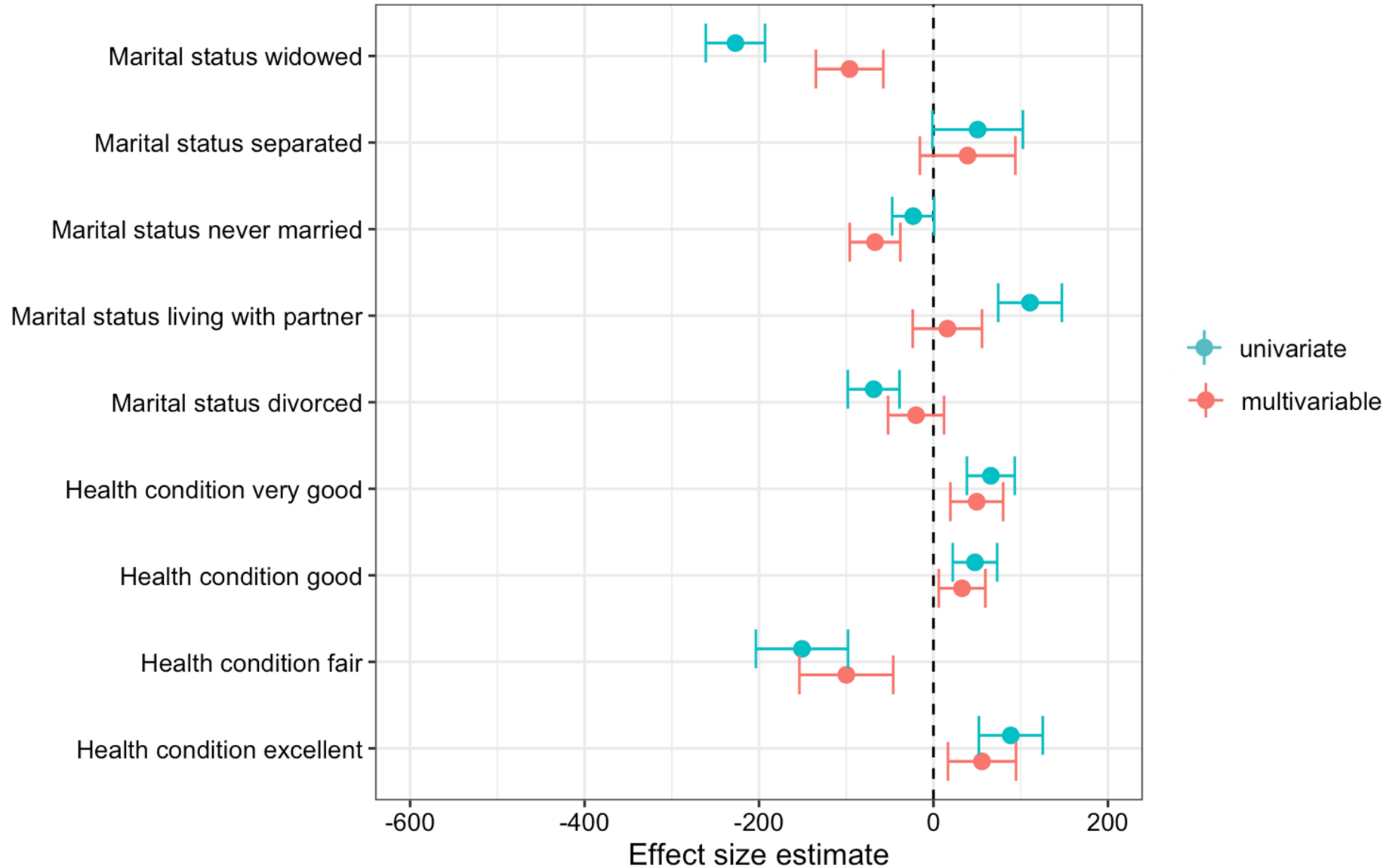
Forest plot of effect size for activity counts

Conclusion

- Accelerometry data is very challenging to work with

  - Complex data structure – there's still a lot to unfold

  - Complex data processing procedure with many steps

- Only medium to low association with TLAC determined despite statistical significance

  - Physical activity volume *might* not be an ideal digital biomarker

  - Further work is warranted

TLAC: Total Log Activity Count

Future Work

➢ Research on alternative physical activity characteristics

➢ Continue exploring the NHANES

➢ Create an R shiny app

➢ Expand functionality of **Arctools** to incorporate more biomarkers

➢ Understand day-to-day variability of physical activity

➢ Develop a framework for using NHANES data as baseline for drug development

Acknowledgements

➤ Project mentors:

  ➤ Debra McIntyre, Exploratory Data Analysis, Regeneron

  ➤ Jacek Urbanek PhD, Digital Health Technologies, Regeneron

➤ Regeneron colleagues:

  ➤ Matthew Wipperman PhD, Precision Medicine

  ➤ Rolando Acosta PhD, Digital Health Technologies

  ➤ Meng Xu PhD, Exploratory Data Analysis

  ➤ Quang Nguyen PhD, Exploratory Data Analysis

➤ Regeneron Biostatistics and Data Management

Selected References

➢ Leroux, Andrew, Junrui Di, Ekaterina Smirnova, Elizabeth J. Mcguffey, Quy Cao, Elham Bayatmokhtari, Lucia Tabacu, Vadim Zipunnikov, *Jacek K. Urbanek*, and Ciprian Crainiceanu. "Organizing and analyzing the activity data in NHANES." Statistics in biosciences 11, no. 2 (2019): 262-287.

➢ Karas, Marta, Jiawei Bai, Marcin Strączkiewicz, Jaroslaw Harezlak, Nancy W. Glynn, Tamara Harris, Vadim Zipunnikov, Ciprian Crainiceanu, and *Jacek K. Urbanek*. "Accelerometry data in health research: challenges and opportunities." Statistics in biosciences 11, no. 2 (2019): 210-237.

➢ Karas, Marta, John Muschelli, Andrew Leroux, *Jacek K. Urbanek*, Amal A. Wanigatunga, Jiawei Bai, Ciprian M. Crainiceanu, and Jennifer A. Schrack. "Comparison of Accelerometry-Based Measures of Physical Activity: Retrospective Observational Data Analysis Study." JMIR mHealth and uHealth 10, no. 7 (2022): e38077.

➢ Smirnova, Ekaterina, Andrew Leroux, Quy Cao, Lucia Tabacu, Vadim Zipunnikov, Ciprian Crainiceanu, and *Jacek K. Urbanek*. "The predictive performance of objective measures of physical activity derived from accelerometry data for 5-year all-cause mortality in older adults: National Health and Nutritional Examination Survey 2003–2006." The Journals of Gerontology: Series A 75, no. 9 (2020): 1779-1785.

➢ Leroux, Andrew, Shiyao Xu, Prosenjit Kundu, John Muschelli, Ekaterina Smirnova, Nilanjan Chatterjee, and Ciprian Crainiceanu. "Quantifying the predictive performance of objectively measured physical activity on mortality in the UK Biobank." The Journals of Gerontology: Series A 76, no. 8 (2021): 1486-1494.

Thank you

Question?