

Digital Health

Opportunities and Challenges

Big Data Course 2016

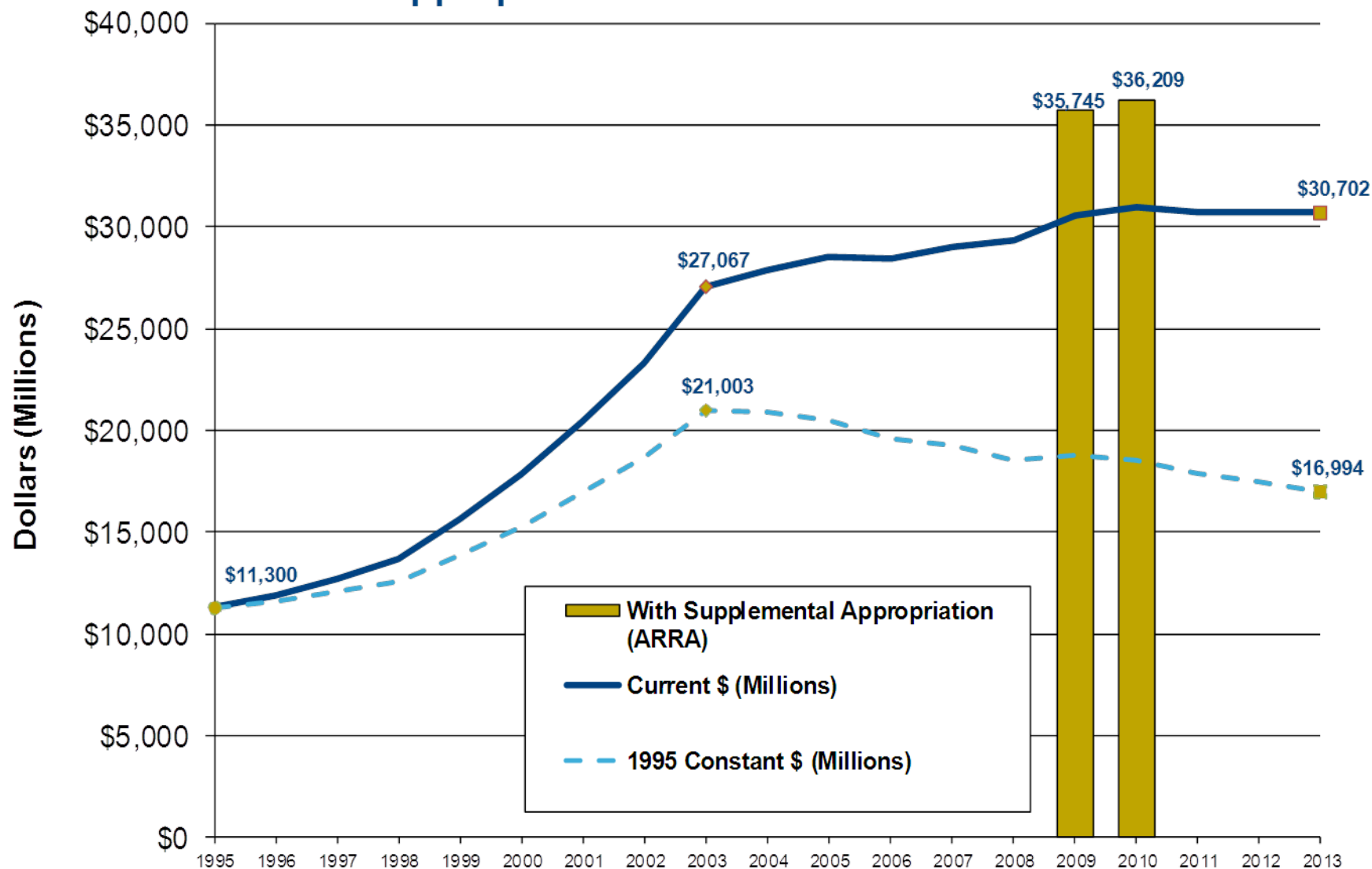
Mark J. Pletcher, MD MPH

Department of Epidemiology and Biostatistics

Outline

- Rationale: Emerging technology → health?
- Health eHeart Study
 - Features
 - Recruitment
 - Data collection
- A (shallow) dive into some big data
 - Step counts by the minute
- Preview of coming attractions in Digital health
- Time for questions

NIH Appropriation in Current and Constant Dollars



EPIDEMIOLOGY

Budget Woes Threaten Long-Term Heart Studies

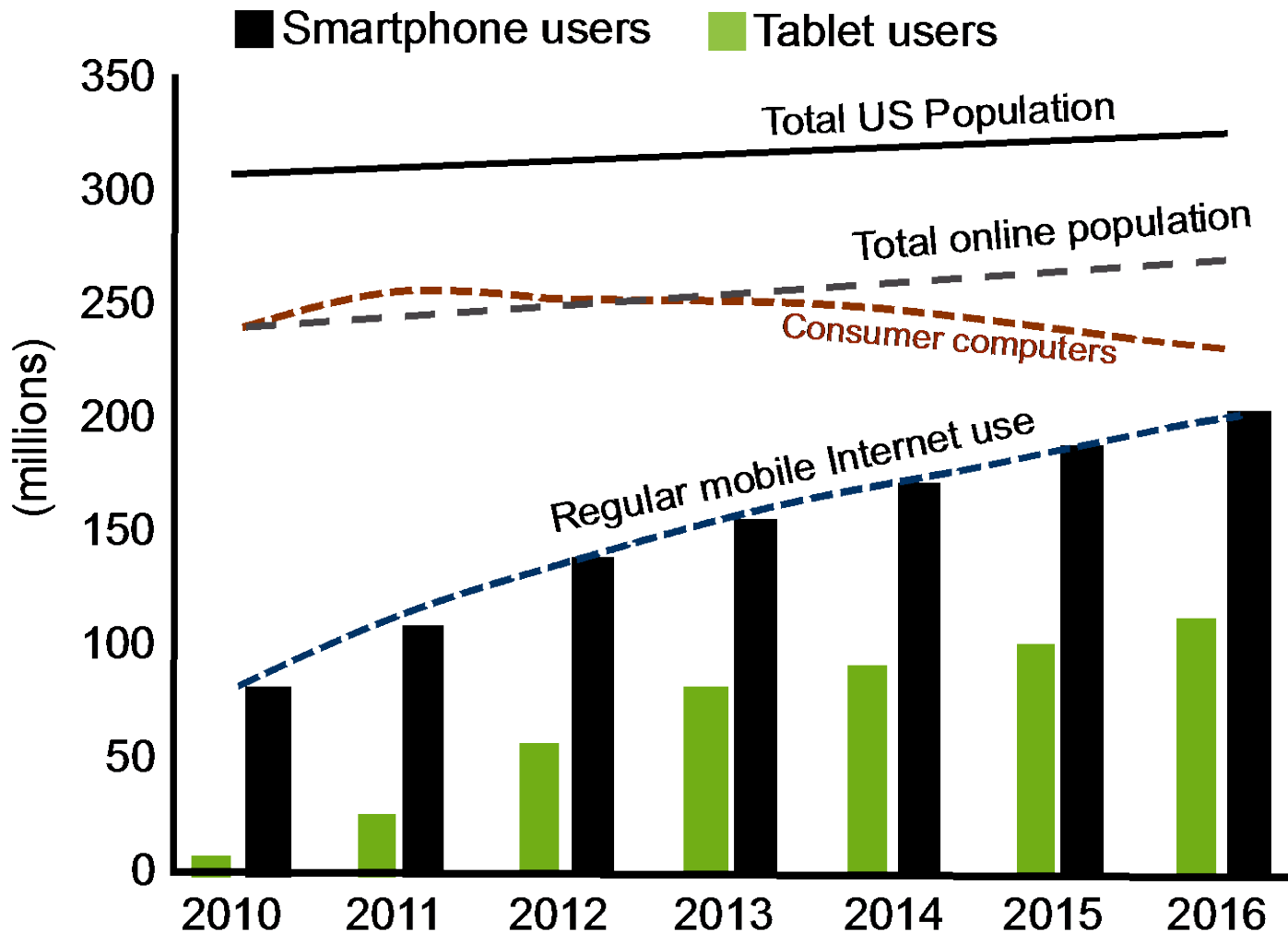
apparent loss in participation, Lauer notes. “We’ve been down this road before and we’ve seen that it’s possible to do this and support terrific science,” he says.

But other epidemiologists argue that there’s no substitute for in-person exams. “Some things could be done more cheaply;

www.sciencemag.org SCIENCE VOL 341 16 AUGUST 2013

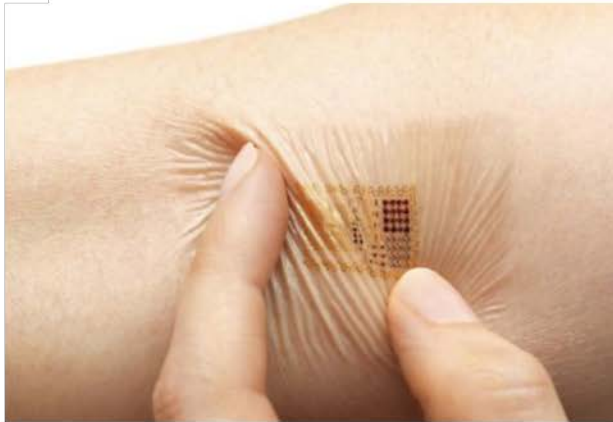
- Framingham Heart Study Budget Cut 40%
- Others affected “even more severely” per NHLBI

US Technology use



Source: Forrester Research

Wearable sensors

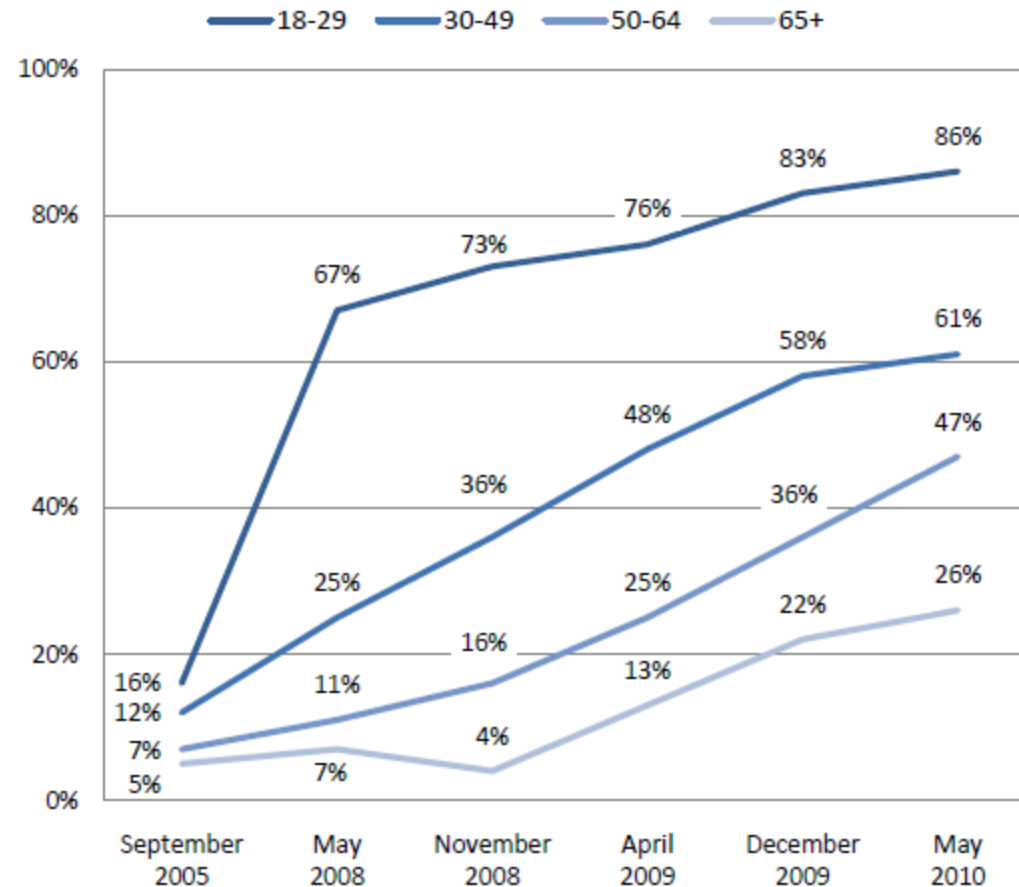


Online social networks

- 1.3 billion Facebook users
- 50% log in on any given day

Social networking use continues to grow among older users

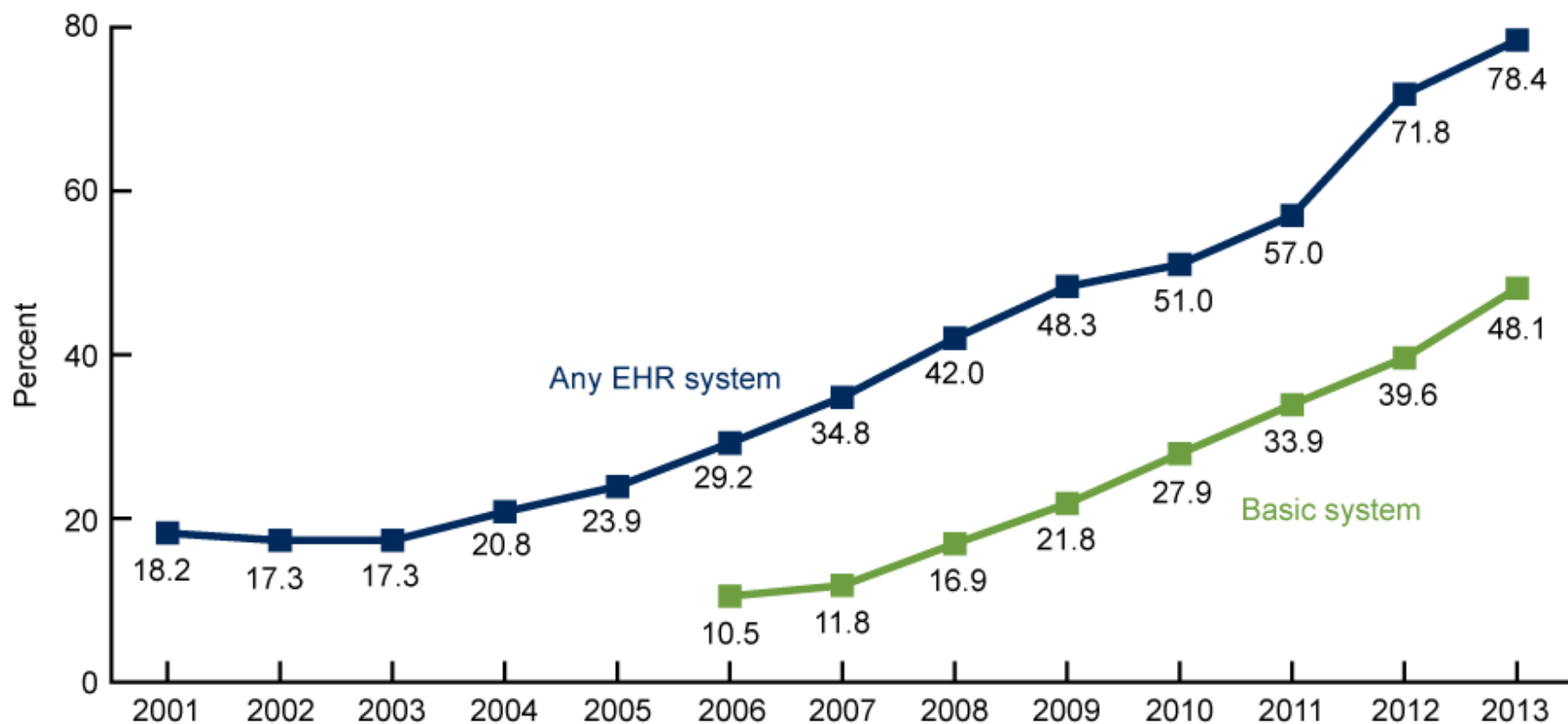
The percentage of adult internet users who use social networking sites in each age group



Source: Pew Research Center's Internet & American Life Project Surveys, September 2005 - May, 2010. All surveys are of adults 18 and older.

Electronic health records

Figure 1. Percentage of office-based physicians with EHR systems: United States, 2001–2013



NOTES: EHR is electronic health record. "Any EHR system" is a medical or health record system that is either all or partially electronic (excluding systems solely for billing). Data for 2001–2007 are from in-person National Ambulatory Medical Care Survey (NAMCS) interviews. Data for 2008–2010 are from combined files (in-person NAMCS and mail survey). Estimates for 2011–2013 data are based on the mail survey only. Estimates for a basic system prior to 2006 could not be computed because some items were not collected in the survey. Data include nonfederal, office-based physicians and exclude radiologists, anesthesiologists, and pathologists.

SOURCE: CDC/NCHS, National Ambulatory Medical Care Survey and National Ambulatory Medical Care Survey, Electronic Health Records Survey.

The Big Question

Can we use emerging technology to improve health?

The Big Question

Can we use emerging technology to improve health?

- 3 Mechanisms:
 - Do research more efficiency
 - Collect new types of measurements
 - Design and test new types of interventions

The “eCohort” Concept

- eCohort $\sim$ Electronic Cohort Study
 - Approach, consent, enroll over the internet
 - Use online surveys for self-reported data
 - Use electronic health records for health measurements and outcomes
 - Use sensors to collect and transmit real-time/real-life data
 - Use online social networks to collect social data, engage ppts, deliver interventions

The “eCohort” Concept

- eCohort ~ = Electronic Cohort Study
 - Approach, consent, enroll over the internet
 - Use online social networks for recruitment
 - Use **Cheaper, faster,** measurements
 - Use **easier, better?** /real-life
 - Use online social networks to collect social data, engage ppts, deliver interventions



The Health *e*Heart Study™

Using big data to reduce heart disease

The Health eHeart Study

- Overarching Goal:
 - Do research that improves cardiovascular health
- Approach:
 - Collect “big data”
 - Keep marginal costs low
 - Support ancillary studies, including RCTs

The Health eHeart Study

- Team
 - PIs: Me, Jeff Olgin, Greg Marcus
 - Study staff led by Carol Maguire
 - Cathy Stolitska, Madelaine Faulkner, Todd Parsnick, David Wen, Madelena Ng, other programmers, coordinators, and collaborating investigators and trainees

Welcome

← → ↻

https://www.health-eheartstudy.org

☆ ⓘ ⋮

Apps google - Google Search Search APEX portal MyAccess Box K program PPRN HeH Group: Health eHeart ... Dropbox Dev site Log In » Other bookmarks

 Health eHeart™

[Home](#) [Study](#) [Technology](#) [Team](#) [FAQs](#) [Login](#)

Join the study to end heart disease.

Be a part of an ambitious study to end heart disease. It only takes a few minutes to make a big difference. Anyone can join — whether you have heart disease or not.

Join the Study



Brought to you by  UCSF

teaming up with  American Heart Association
Life is why™

BASIC INFORMATION

Contact Information

Basic Demographics

ACTIVITY & WELL-BEING

DASI Survey

IPAQ

SA-NYHA

SF12

HABITS & LIFESTYLE

Detailed Smoking History

Second Hand Smoke

Alcohol Use

Epworth Sleepiness Scale

Pittsburgh Sleep Quality

FFQ

CONNECTIVITY

Social Demographics

Emergency Contact
Information

Internet/Smartphone Use

Internet Use for Medical
Information

MEDICAL HISTORY

Mother Family History

Father Family History

Sibling Family History

General Family History

Medical Conditions

Symptom Overview

Past Cardiovascular
Procedures

Medications

TRIGGERED

Coronary Disease Survey

Seattle Angina Survey

Heart Failure Survey

KC Cardiomyopathy
Survey

AF Survey

AF Severity Scale Survey

AF Effect on Quality of Life

Chest Pain Symptoms

Shortness of Breath
Symptoms

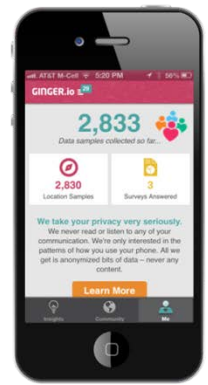
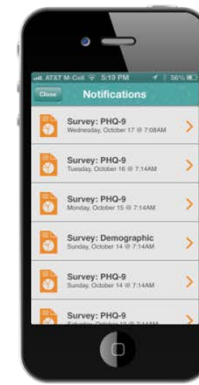
Palpitations Symptoms

Fainting Symptoms

Calgary Syncope
Survey



GINGER.io



- **Background continuous data collection from smartphone sensors**
- **Reminders/Messaging**
- **Contextual alerts**
- Mobility & Activity
- Screen interaction, call/text interaction
- Behavior modeling

iHealth™



Pulse Oximetry



BP



Wt/Body Comp

fitbit



Integrated devices and apps

Withings



Modular Consent System

- Start with “umbrella” consent to answer surveys and use data for research
- Then offer easy, short, just-in-time consent modules for each optional feature
 - Each device company
 - Medical records/HIPAA Authorization
 - Food Frequency Questionnaire
 - Ginger.io and Azumio apps
 - In-person testing, biobanking

2-Step Fitbit Consent Module

Connect your Fitbit

It's easy. Clicking the button will take you to their website to enter your Fitbit login and authorize the connection.

Once you do this, we can import the data from you Fitbit devices into your profile, and we'll use it to fight heart disease!

We'll keep your Fitbit data (private and secure), just like we do with all your study data.



Connect

Cancel





Health eHeart Local by UCSF would like to access your profile and data on Fitbit.com:

60 + READ AND UPDATE ACCESS

Email

Password

[FORGOT PASSWORD?](#)

ALLOW

You can revoke this application's access from your Fitbit account settings.

Not using Fitbit yet?



-  TRACK ACTIVITY & SLEEP
-  MANAGE WEIGHT
-  WATCH YOUR PROGRESS

Fitbit motivates you to turn fitness into lifestyle, challenge friends, set goals, and watch your progress.
[Learn more about Fitbit](#)

[CREATE A NEW ACCOUNT](#)

Modular Consent System

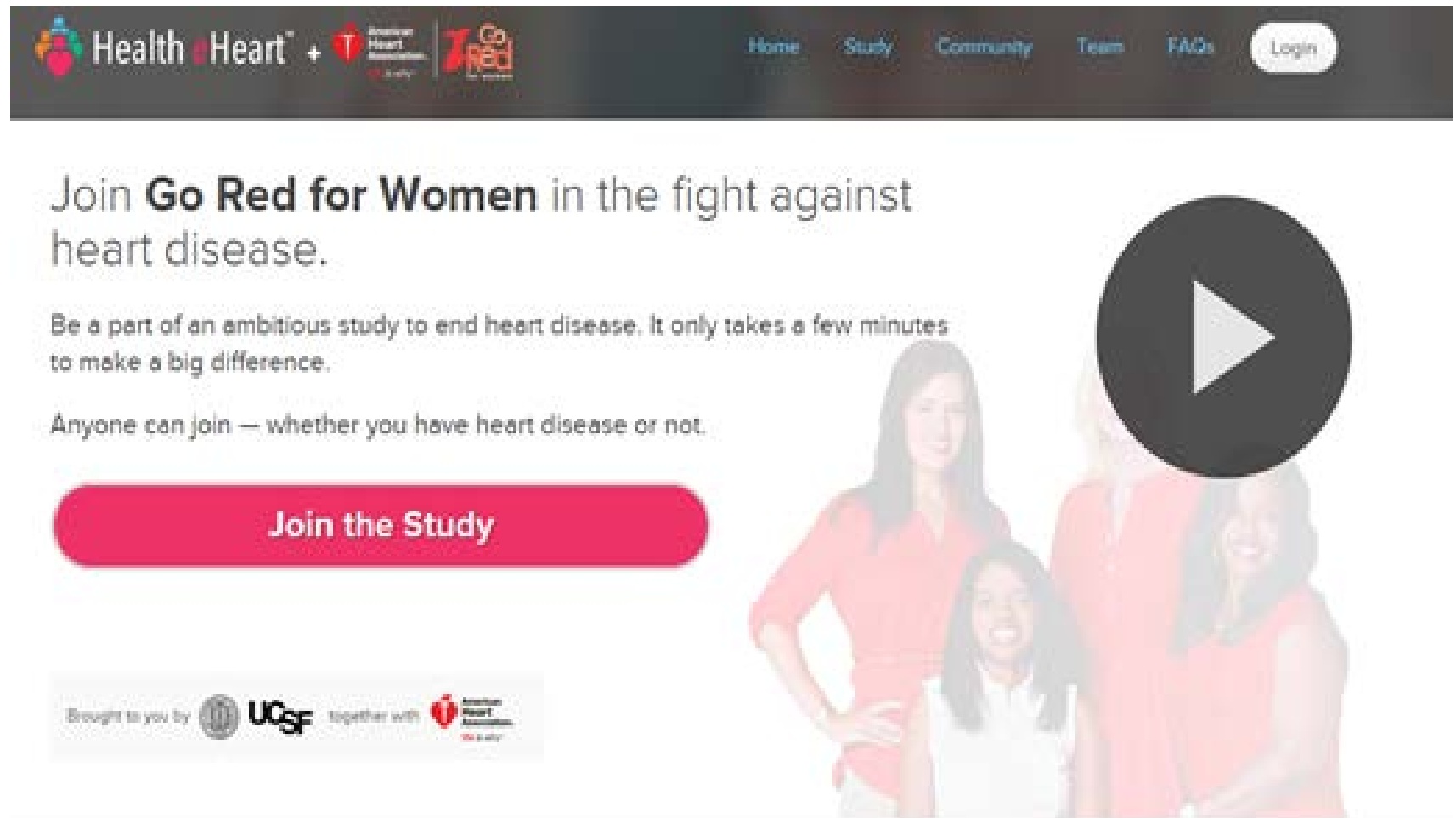
- Tech Detail: We use Oauth 2.0 – an industry standard – for authentication/ “integration”
- Extremely efficient and secure

Referral and Co-Enrollment Tracking Systems

- Unique URL for each referring partner
 - Customized landing page
 - Track source of recruitment
- Add identifier for “co-enrollment” of individual participants
 - <http://www.health-eheartstudy.org?rfk=4b62fda3607d1425482271&id=ABC123>
 - Link with data assets provided by a referring partner
 - “Mail-merge” is only requirement for partner

Example #1

<https://www.health-eheartstudy.org/gored>



The screenshot shows the top navigation bar of the website with the Health eHeart + American Heart Association logo on the left and links for Home, Study, Community, Team, FAQs, and a Login button on the right. The main content area features a large heading, a subheading, and a call-to-action button. A large play button icon is overlaid on the right side of the page. The footer includes logos for UCSF and the American Heart Association.

Health eHeart™ + American Heart Association

Home Study Community Team FAQs Login

Join **Go Red for Women** in the fight against heart disease.

Be a part of an ambitious study to end heart disease. It only takes a few minutes to make a big difference.

Anyone can join — whether you have heart disease or not.

[Join the Study](#)

Brought to you by  UCSF together with 

Example #2

<https://www.health-eheartstudy.org/?rfk=69da54e535641439234236&id=A>

 Health eHeart

[Home](#) [Study](#) [Community](#) [Team](#) [FAQs](#) [Login](#)

 UCSF Medical Center

Welcome UCSF patients! Join the study to end heart disease.

Be a part of an ambitious study to end heart disease. It only takes a few minutes to make a big difference. Anyone can join — whether you have heart disease or not.

[Join the Study](#)



Brought to you by  UCSF in collaboration with  American Heart Association

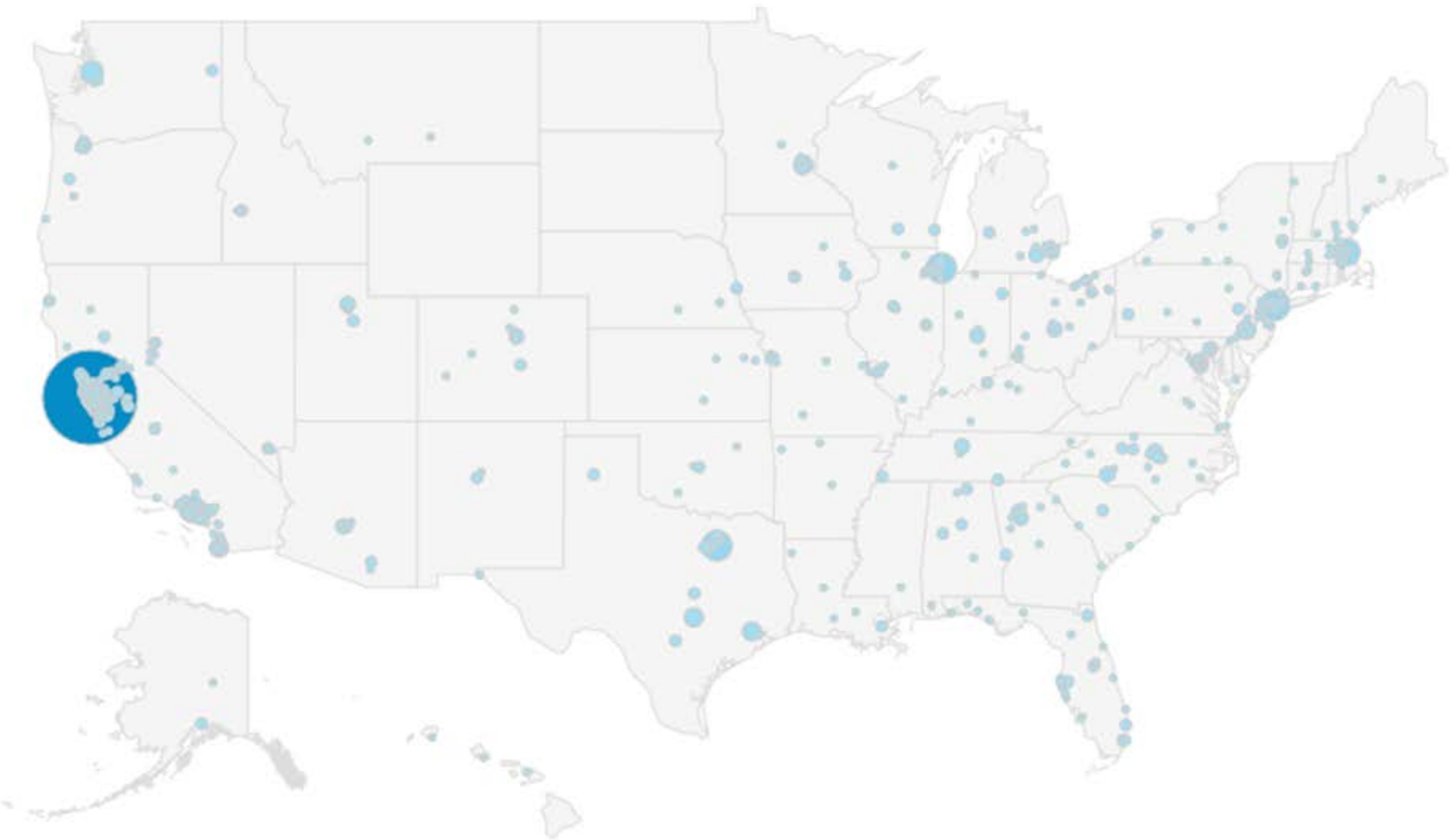
Referral and Co-Enrollment Tracking Systems

- Successful campaigns
 - AHA's Go Red for Women
 - → 11,000 participants (almost all women!)
 - UCSF patients
 - 200,000 emails → 6,000 ppts (3% uptake)
 - Linkable to medical records if they consent
 - Childhood Cancer Survivors Study
 - 500 → 310 (63% uptake)
 - Linked Fitbit data to CCSS Cohort data

Recruitment so far



United States Geographic Distribution (Cities): Consented Participants via Google Analytics



1 962

Digital Health Volunteers (non-random)

<u>US pop</u>		<u>Health eHeart</u>		
NHANES		All	eVisit done + Fitbit	
<u>n=6,113</u>		<u>n=26,982</u>	<u>n=16,501</u>	<u>n=1,276</u>
Age, mean	47	51	52	49
% Female	52%	75%	74%	68%
% AA	12%	6%	4%	2%
% College+	29%		71%	74%

Measurements collected

As of 8/9/16

Self report

- 815,928 surveys completed (from 33,663 people)
- 146,207 vital signs (from 20,764)
- 59,090 lab results (from 8,976)

Participant-triggered device measurements

- 506,145 weights (from 1,372)
- 108,437 blood pressures (from 823)
- 18,657 EKGs from AliveCor (from xxx)

Passively collected device/app measurements

- 2,024,354 daily step counts (from 6,454)
- 314,361 days of movement/communication patterns (from 2,265)
- 2,429 possible hospitalizations from 179 hospitals

Measurements collected

As of 8/9/16

<u>n/N</u>	Self report
24	• 815,928 surveys completed (from 33,663 people)
7.0	• 146,207 vital signs (from 20,764)
6.6	• 59,090 lab results (from 8,976)
	Participant-triggered device measurements
368	• 506,145 weights (from 1,372)
132	• 108,437 blood pressures (from 823)
	• 18,657 EKGs from Alivecor (from xxx)
	Passively collected device/app measurements
314	• 2,024,354 daily step counts (from 6,454)
139	• 314,361 days of movement/communication patterns (from 2,265)
	• 2,429 possible hospitalizations from 179 hospitals

Measurements collected

As of 8/9/16

n/N

Self report

- 1.4 • 27,554 weights (from 19,359)
- 1.8 • 25,798 blood pressures (from 14,133)

Participant-triggered device measurements

- 368 • 506,145 weights (from 1,372)
- 132 • 108,437 blood pressures (from 823)

Measurements collected

As of 8/9/16

n/N

Self report

- 1.4 • 27,554 weights (from 19,359)
- 1.8 • 25,798 blood pressures (from 14,133)

Participant-triggered device measurements

- 368 • 506,145 weights (from 1,372)
- 132 • 108,437 blood pressures (from 823)

Much more efficient and less error-prone to get measurements directly from a device...

Typical digital data flow

- Sensor → Smartphone → 3rd party server → Health eHeart server (JSON) → Extract/Transform/Load to .csv files → Statistical software on Analyst's computer system
- JSON = "JavaScript Object Notation"
 - JSON is a data format, stores data in key-value pairs
 - {"key1":"value1", "key2":"value2", ... , "keyN", "valueN"}
 - Values are encoded and compressed

cpub-QuickSessionCollection-QuickSessionCollection-CmsRdsh - 654MNRDS-1.PSG.NET - Remote De...

user_id	diastolic_value	systolic_value	heart_rate_value	source	date	time
12 80 120 77	manual	02/20/2015	18:55:00.000			
12 80 120 70	manual	12/16/2013	0:45:00.000			
12 80 120 70	manual	12/17/2013	0:45:00.000			
12 80 120 77	manual	02/20/2015	18:55:00.000			
16 73 138 55	device	02/21/2013	21:52:00.000			
16 74 130 56	device	02/22/2013	16:40:00.000			
16 69 123 55	device	02/22/2013	16:36:00.000			
16 72 128 56	device	03/07/2013	14:51:00.000			
16 72 126 58	device	02/22/2013	16:38:00.000			
16 72 123 53	device	02/23/2013	22:06:00.000			
16 72 128 56	device	03/07/2013	14:51:00.000			
16 67 130 54	device	02/23/2013	22:05:00.000			
16 101 163 85	device	05/24/2013	19:05:00.000			
16 71 127 50	device	02/23/2013	2:05:00.000			
16 82 131 67	manual	02/21/2013	21:30:00.000			
16 72 124 53	device	03/07/2013	14:51:00.000			
16 71 116 52	device	02/22/2013	16:41:00.000			
16 70 128 47	device	02/23/2013	2:06:00.000			
16 70 125 47	device	02/23/2013	2:07:00.000			
16 68 119 53	device	02/23/2013	22:07:00.000			
16 71 127 53	device	02/22/2013	16:39:00.000			
16 72 124 53	device	03/07/2013	14:51:00.000			
66 73 112 70	device	07/31/2014	19:33:00.000			
66 73 120 62	device	10/13/2013	6:44:00.000			
66 69 109 67	device	10/12/2013	16:47:00.000			
66 71 114 65	device	10/06/2013	15:30:00.000			
66 57 99 70	device	10/14/2013	5:37:00.000			
66 65 109 60	device	10/18/2013	6:13:00.000			
66 69 115 74	device	04/10/2014	14:06:00.000			
66 58 98 74	device	10/10/2013	15:42:00.000			
66 67 115 60	device	10/18/2013	6:15:00.000			
66 67 113 73	device	10/09/2013	14:59:00.000			

.CSV file of
Blood pressure
data from
Withings cuffs



cpub-QuickSessionCollection-QuickSessionCollection-CmsRdsh ...



meas_fitbit_tracker - N cpub-QuickSessionCollection-QuickSe

File Edit Format View Help

user_id	date	steps	distance	calories	floors	elevation	device_type
12	10/25/2013	0	0	2060	0	0	
12	08/07/2013	0	0	2072	0	0	
12	01/17/2015	0	0	2146	0	0	
12	11/20/2014	4912	2.0510593128	2034	0	0	
12	06/30/2014	7029	2.9502281664	2140	0	0	
12	01/15/2014	10404	4.3443104752	2408	0	0	
12	04/04/2013	10123	4.3752485468	2118	16	160	
12	09/23/2014	8534	3.5634705804	2252	0	0	
12	01/26/2015	10662	4.4520375987	2408	0	0	
12	10/31/2014	8233	3.4377858294	2208	0	0	
12	08/22/2014	7244	3.0248163215	2150	0	0	
12	10/28/2014	7447	3.1096086344	2104	0	0	
12	02/06/2014	9190	3.837389643	2194	0	0	
12	06/05/2014	0	0	2086	0	0	
12	02/06/2014	5000	3.4444444444	2000	0	0	

.CSV file of
Fitbit activity
data

One JSON “packet” from Fitbit: encodes 1440 different by-minute measurements...

```
{"calories": "eJzNVy1vFVEQTRp+AD+h+wOAP9CURT6BJBhShcNV1uCaIEAhCK7BoHCAIUeqSHCIR/gDTzTBgSKpgHMfZ+/Z2Zm5+9oUnpjs7t2783nmzN3l8+Oby0CeHZ+E784rl6Fz2+SiMeL7OTr+Vy4zu3N9v6ycbLO0/Mf7o7PrPa73Dvf6TeLIXs1T9L2uqy0VrNGX/dNFj+d/IYdtqcd5xcvVOpfdkFP7bvfg7oXzq/WjPfgC3XhWP3CPunr1xx6t/1g6+02/s7w/sqMYpqwWhz3sYS/udW1qY4056mrFzL2KVbUPmx7ecdX9tjaM1fYF4oTvJ2697B+8fd2/ufq0XI/OXv3R9a7c47p/+rm/fe0DbJdnvclv0IM9WheLCdvPtq+9NX1HTK190lp2oxqJjuKj2v9453GJE3qYS8ZJ35kT7sUaYmReP71/OOioPtWYFVNZjNzj43x6bzmOV8autR7b9e2M16q/vq9je9HeqM5T3pj3he/be5iXy1+rAXdY0uNV67UQ21L/EN9rDmihViQjGGNevLNIZpHhS7Hudltcvy7uW5zbPdWBfkAvAgY1cOwxp7PtLv1SvDFnSy7zwsIsfo5RsvvhTBflzJk3gGZ3n8ZufY1yu/igAP0BvVjXX2chvNKc4Zv9ZTPHi9kGGNlfOsn7L+Yt7/znl3Ttq4rP9ZX0e+e1g3MsziqAZzauRxExDDWW3nB/u9+tRNZlr1sUv7rJWbLF8Z1qzoOSj2tV2TzLbhvIR3hl2LGZt7b5YpBiOdeezT3OCKmMAFO8vVcMax9okTXInVCGMe/yp+aRN29ByX5T6qUZaHCLfgScS6WvwoAi7kPYTvfj757go4k2ehet6tc6yeXf0zxS8mmFF56a3z7MX2deeZ629fYxtWc6WewMmdg++lvzgbLz8gLCiQtAti018h/M3vrVnAs5HX1zeHDBoscd5GfVpi1db51PloRbfbMJrXde38/hTZ+jbU673p7FNK+WL1o8Ogf/xJtiMhLtN493lpzMmUt2D2cle8Ge2WhX/2H1W+X38XP9Dx+fB32fal66wQ+eS/U/Tv+Hsz7KuLc151R+AxOMThs=", "steps": "eJztVWtPE0EUPQLn9vu0tluAulDgTY+oiZg1Ch+al0mVglxGAUPwr4wPD/E87M7OzO7KtbwyfjaWa7c3fm3nPP3JkB/uNfgMvW/auZLeOtwWcf/vVQumYsRG81VFCNeg2slvTQhIM2luXcOub4XDL8uujQqLGB0nOMe5yzwjnim0+bj/vUaAMDvg/QM/z3GKGOJAK5NgG9i/ktrid20W/R7kk2cVQFD2t4gHfYwwj7mOCAbwc4xhGe8/8DWQEFyatWMneBKsdPGFPMcdhrhvYdelR4wdzWyMyn7wmVcqU+r7DFOXck10dkdZu9m/lbpKckTEsJJzmEse3M8+3xP7ykffdMXik2SjrY8bcYNtidjWuidL2LrVeohYLVGWXKy6wTU3SK21yd61+WiM7ch7jmmyCMddN7ck+eaxnjM3ycAuuiiLlu/iEr/jC1TthPk3umlFKYw+X/D3FG8s6LGCefVZMW6XpmGedzlpZHSO3+AQoWkfTh8YYm9zFZbGMRYjzQPnvklNqydYqXZ3T2KThTqmLNB6WHCf0V3zaVp12CmeZbAZ4glNWzHG0S4UmAflz176H+MyzYjbEPkc4wwV+swr/8HmB7/iFc/xk+yHbS3kKDFmsyCwDeaqXQz3kap6ZRavWZaTastgzGeyDvv4xnPikLvqbWKWvk8dvMZ7Vp6OsJKTsUBD3gKKjz0uCTccr1E++9mhomj/TsRZ7cNZ6r1PLYOErSo9Ky++FQkZt5q21Jlxj71WxKbLytR6VKljU47qsqk7fBXmySGyanNGhd89eYsKBh4tgkOx9lm4AnuwJww=", "distance": "eJztmE1IVFEUx9WohhRcJFKkVCluajVEFr57GQMJB2kTbbRVCOqiKISciKSEqCB00QcqtZCEMDd9IAYGA4WBC6EwKNrMooUuxHATGC66v/c44/N5Z3zzYbSYA8Odue++e8/5n//5uFNSUpSiFOV/kZromMPYuFaq/tWZyeRrxY3NH/Pne/or1YNSz82rSmKXRZHFjd1bDU4dRE653RgdimZ+B5sgvq747+J47xWdPoQMJJQnKzp2IXzN6jDX3eijbPrCwYn4b4e1wX2xgVH4ObxrSZ1inOap06658h7clZ1jOPVDxU8qjh8Uxl7FfOMqy8rrXYzjw02HYly3V6rJG74ju5yPjyO917eciY85xnrO/VcChubrX5hbac+g26coxK1u7XBT125v0dXvtmhW2MVmjnG8uX92tisZ/75oPhtbNZgxT4L3Ts1eMGLrezLRfAD+6MTdso5HidiDtj61zdPvXN19M+VdX1S+A17wQw80dvbM5rCHP5UjX9WC913XHvufrygBNfp9mCuVmbefVbWdTr1ngg6bWVPujWSo8TmMLjY7Bex+TyXNTb9cpFs3/fj5McjDDb+tcZHCjuJV0Z8i+/wIXHi5+3iSi+bA8gn8AI+wJWrt765MS7rBtsmFTWjE9M2QY8gd/wShkfpJBt85BwwkFxAHvTieb1Oku/Ai5h/8tVhpuqcux/xZbMZ3E3+1BdnDulHIQc1sddyo07jH3hPrSFncvJ9j/yqkXz6YtMYFtaO1pjD7KuX9ibTV+RbywVQobml47XG0UceE7etPnN1gPkGo/p9AjOwcHGtdtuLc5nb5vAY3Om62NiOK8/tQFMhE+ZeBUmd+aLTRjBhnzyRRgpXx4uaE8n/N/lo0TaHkn6nFwkQ71Vc09fqKnHxzQ5xutxNtZS4Qk9AGO2Pg/qTW75+fyApq/lxZawYtOTPDnYdlwb/fTsd+3mwustp9zvzJm40Ct1SveUntD7LjW5o6kDqfHtl9IUL4QdXq9an7pb8Vt69ULalpwQXNP1mbnGGv2719/VpvVjX+Qo9d9ZqXuvJB9SE439mn5ivHqvplbFe39l9GvwfgpWM1Vf1fnJVbfuBm2gDqfbKxMX0VXuAiLSR2fSL5P4+xvB/doO3/8L8dsS1J/fQZz99XC78jvcgpeZOCnni85+Xfy9iM0nYe5qwTXEIT7mPsszf50U8e7CERXkB7rBR8mv7Mva7w6/fg+X/w5sPQfi/b+QoGdxcw/vs6/cRQUd5lgjOOTD+7DyF2ZZM3A=", "floors": null, "elevation": null, "device_type": "MobileTrack"}
```

id	user_id	measurement	date	minute	value	resource	device
50108	12	3594043	29MAR2013	567	20	steps	MobileTrack
50109	12	3594043	29MAR2013	589	15	steps	MobileTrack
50110	12	3594043	29MAR2013	592	28	steps	MobileTrack
50111	12	3594043	29MAR2013	621	9	steps	MobileTrack
50112	12	3594043	29MAR2013	656	10	steps	MobileTrack
50113	12	3594043	29MAR2013	659	22	steps	MobileTrack
50114	12	3594043	29MAR2013	660	112	steps	MobileTrack
50115	12	3594043	29MAR2013	661	112	steps	MobileTrack
50116	12	3594043	29MAR2013	662	111	steps	MobileTrack
50117	12	3594043	29MAR2013	663	111	steps	MobileTrack
50118	12	3594043	29MAR2013	664	111	steps	MobileTrack
50119	12	3594043	29MAR2013	665	109	steps	MobileTrack
50120	12	3594043	29MAR2013	666	111	steps	MobileTrack

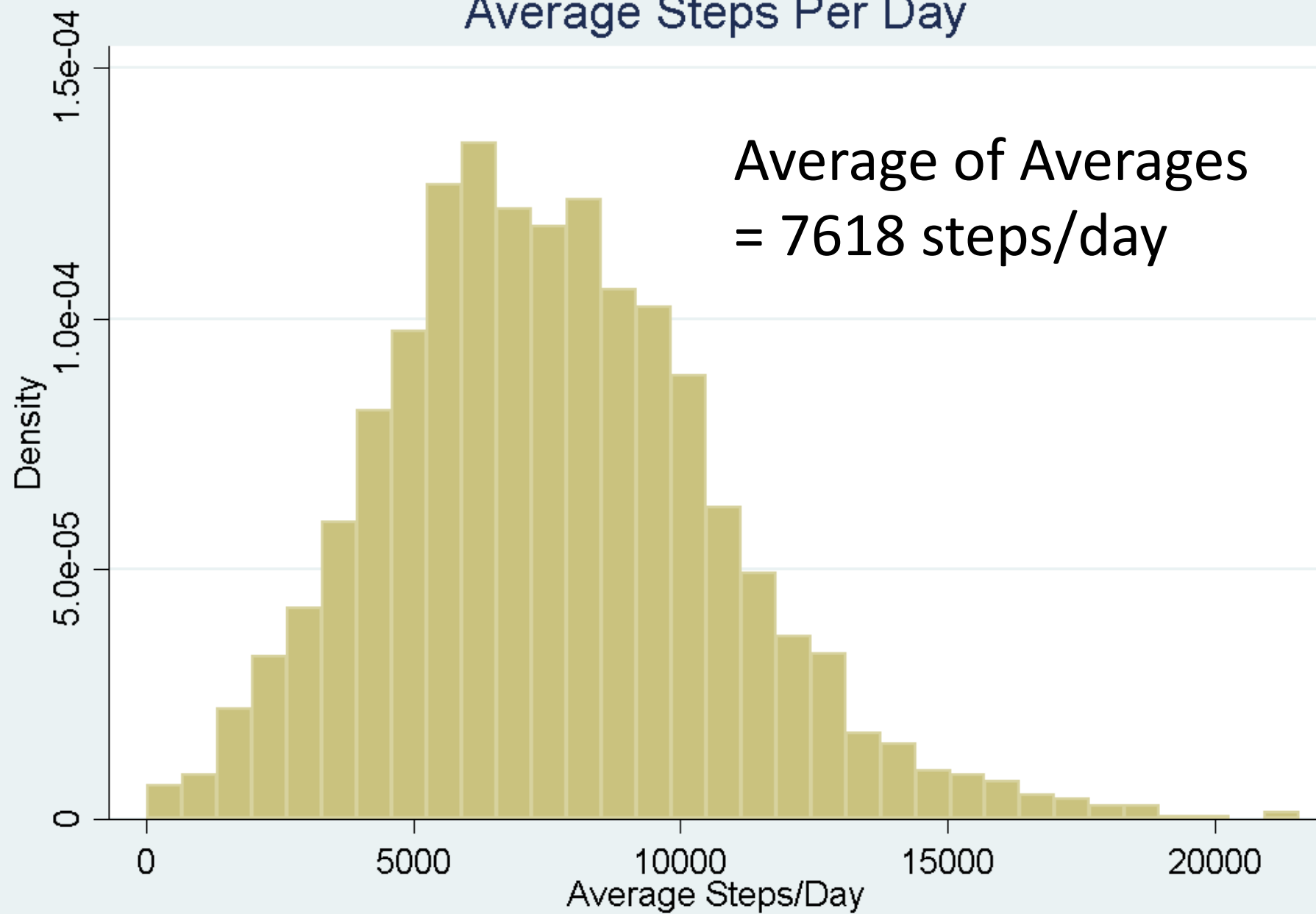
Stata screenshot of Fitbit “Intraday” data after it is unpacked and imported into Stata

Big Data

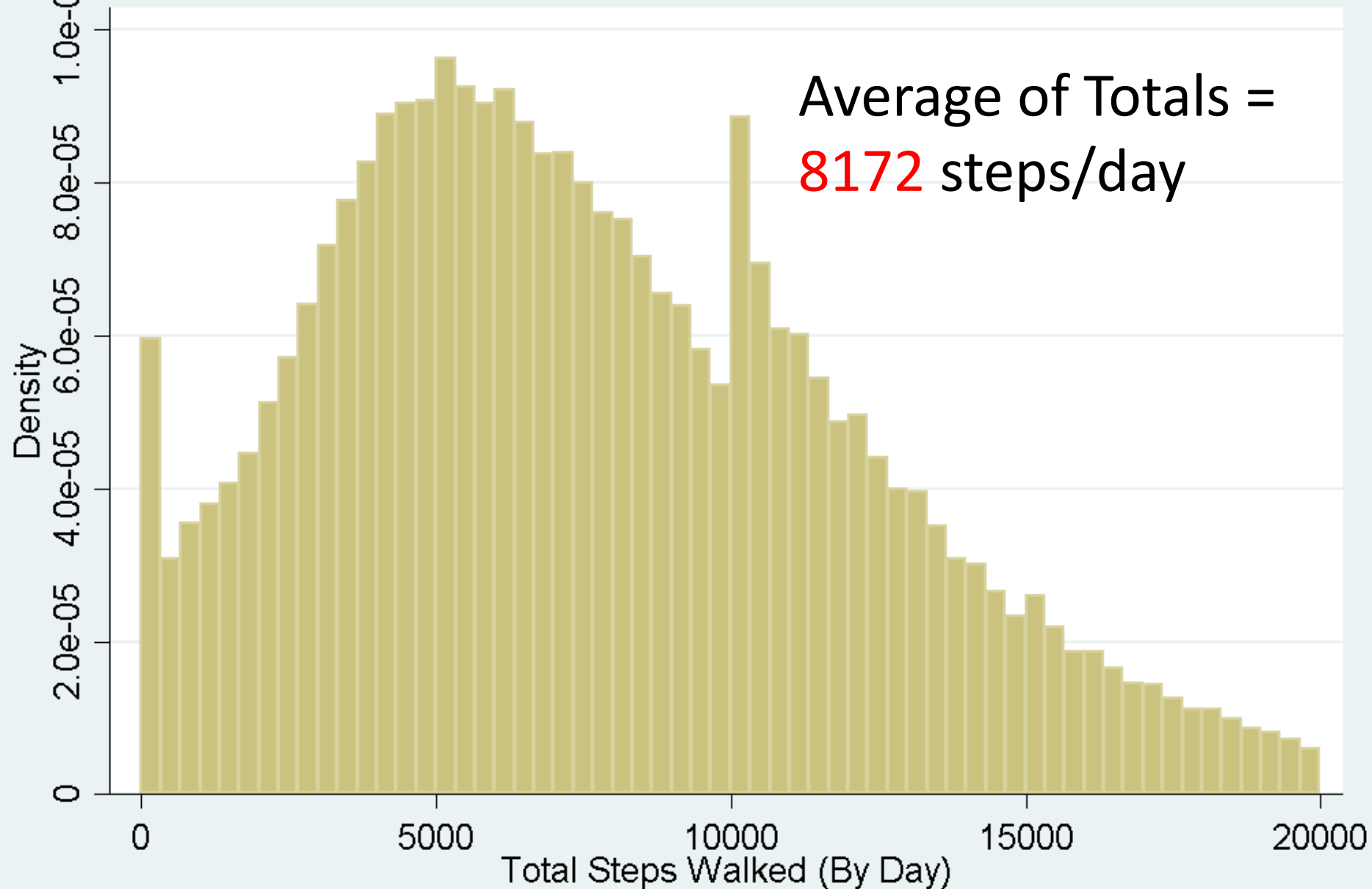
Example: Health eHeart Fitbit data

- Step counts per day – not that big
 - 2213 participants contributing Fitbit step data
 - 1.04 million total days of data (observations)
 - 471 average days of data/ppt (range 1-2356 days)
 - 45 MB
- Step counts per minute – pretty big
 - 231 million total minutes of data (observations)
 - 221 average minutes/day with non-zero step count
 - Compare with 1440 minutes/day (=24 hours/day*60 minutes/hour)
 - 12.1 GB

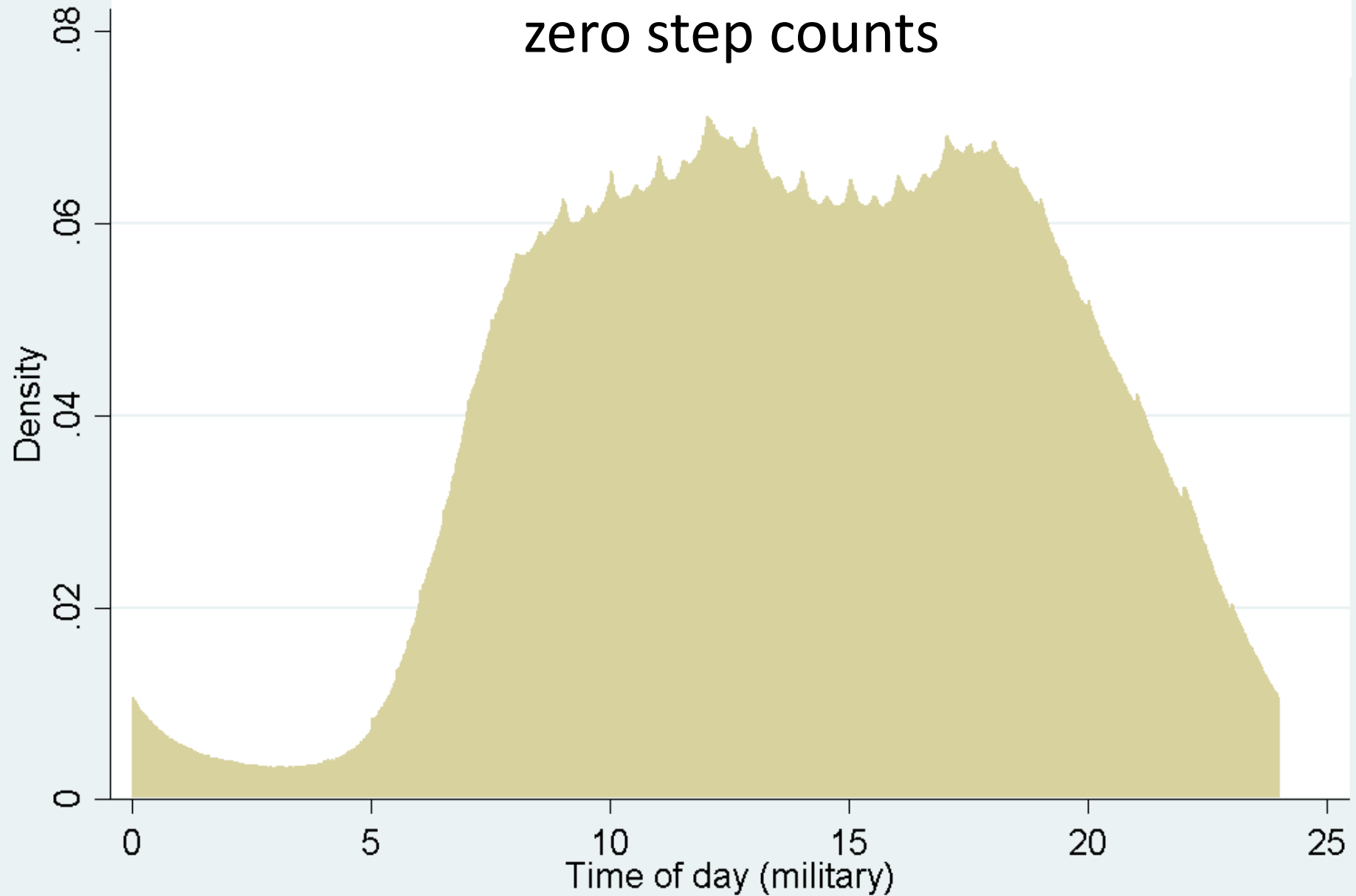
Average Steps Per Day



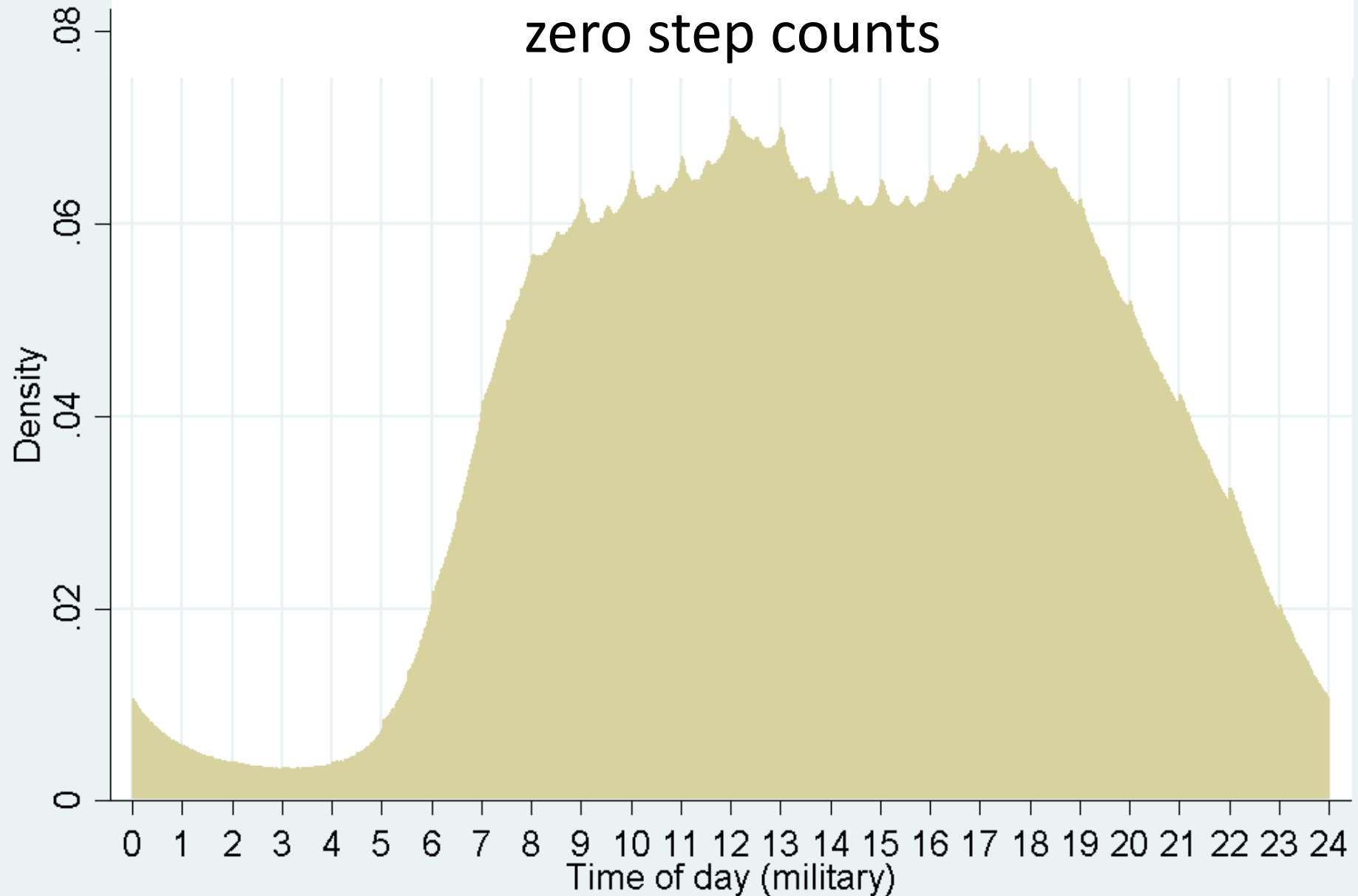
Total Steps Walked Per Day (trim to <20,000)



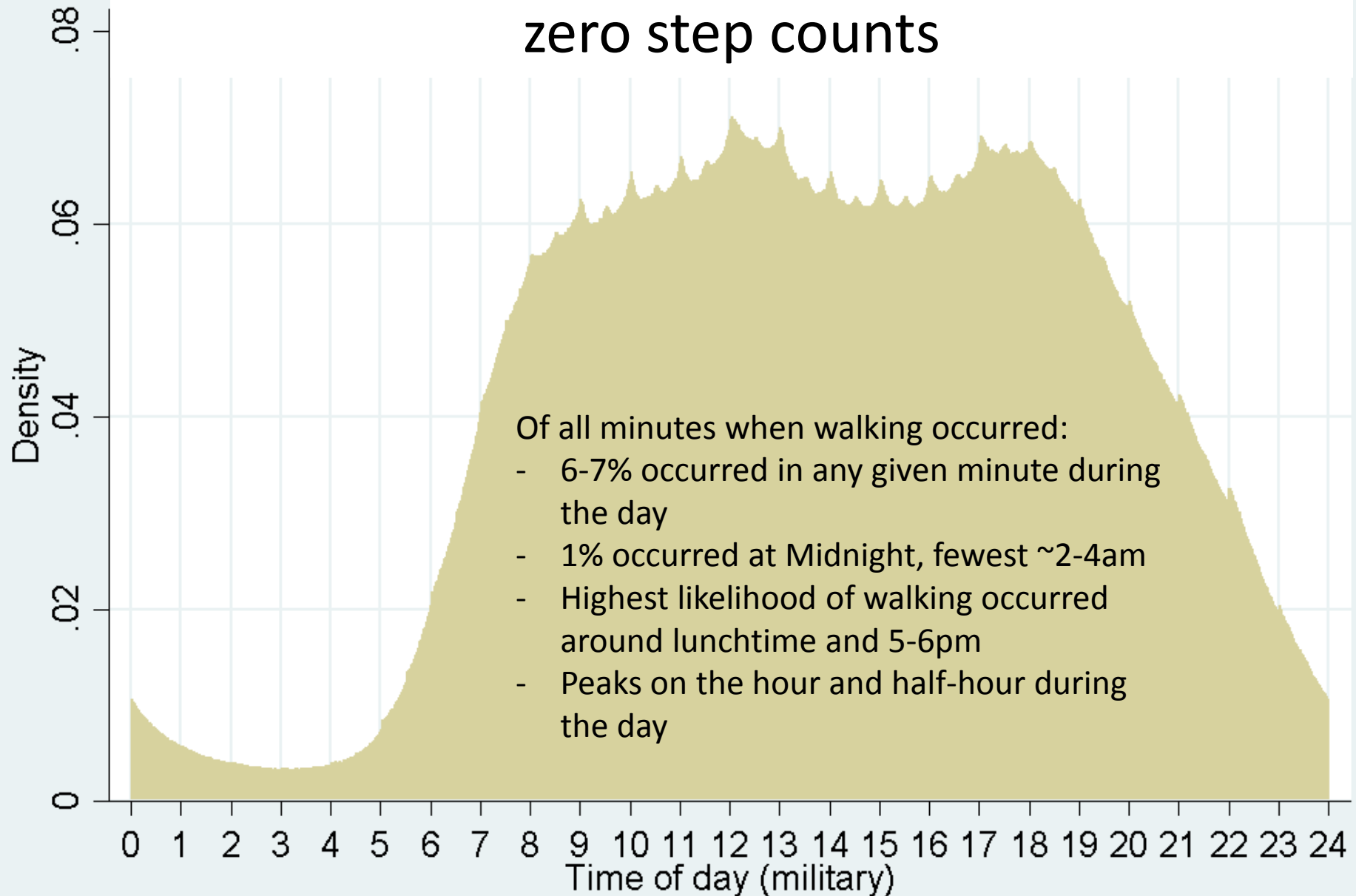
Distribution of minutes in the day with non-zero step counts



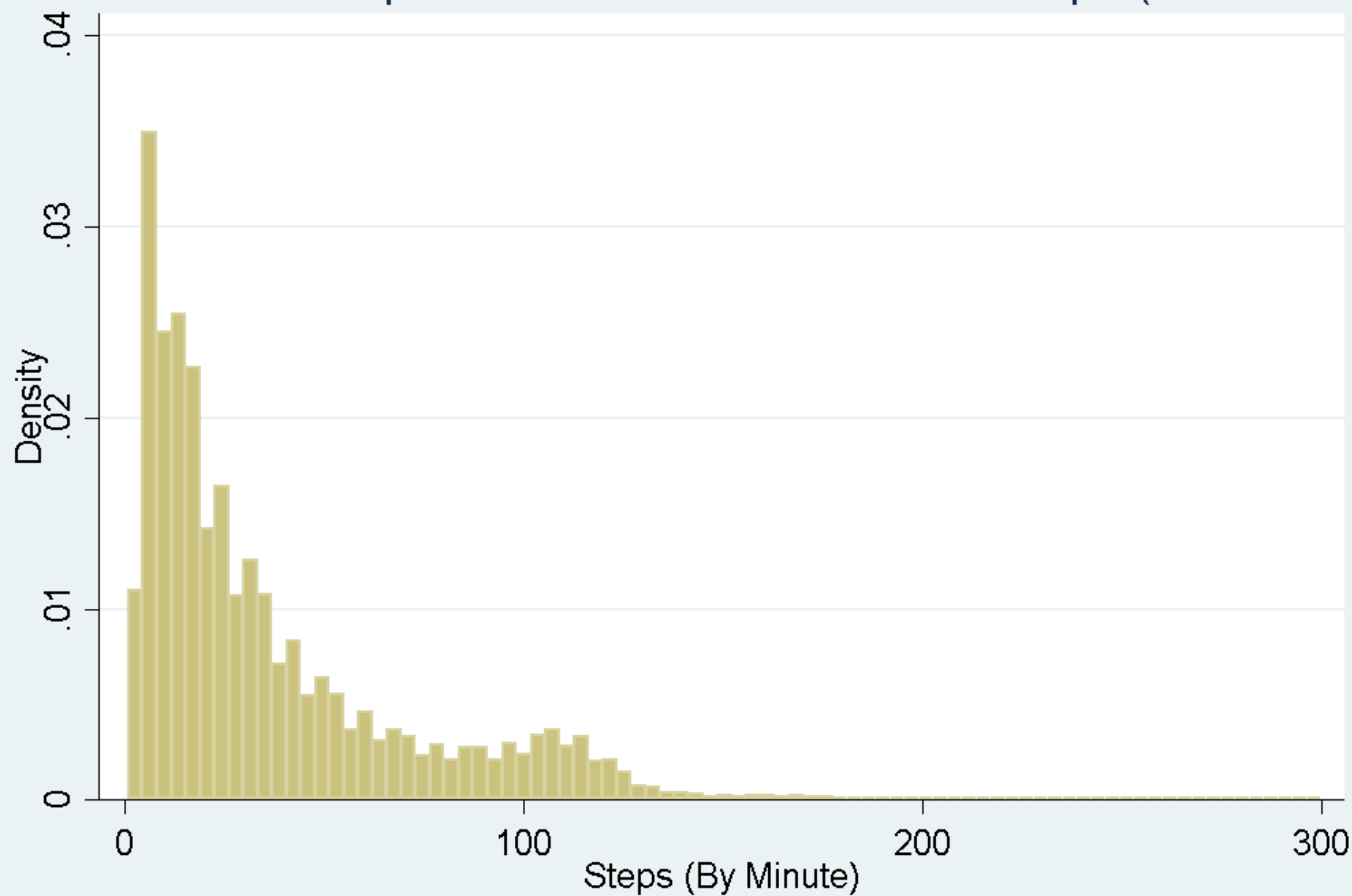
Distribution of minutes in the day with non-zero step counts



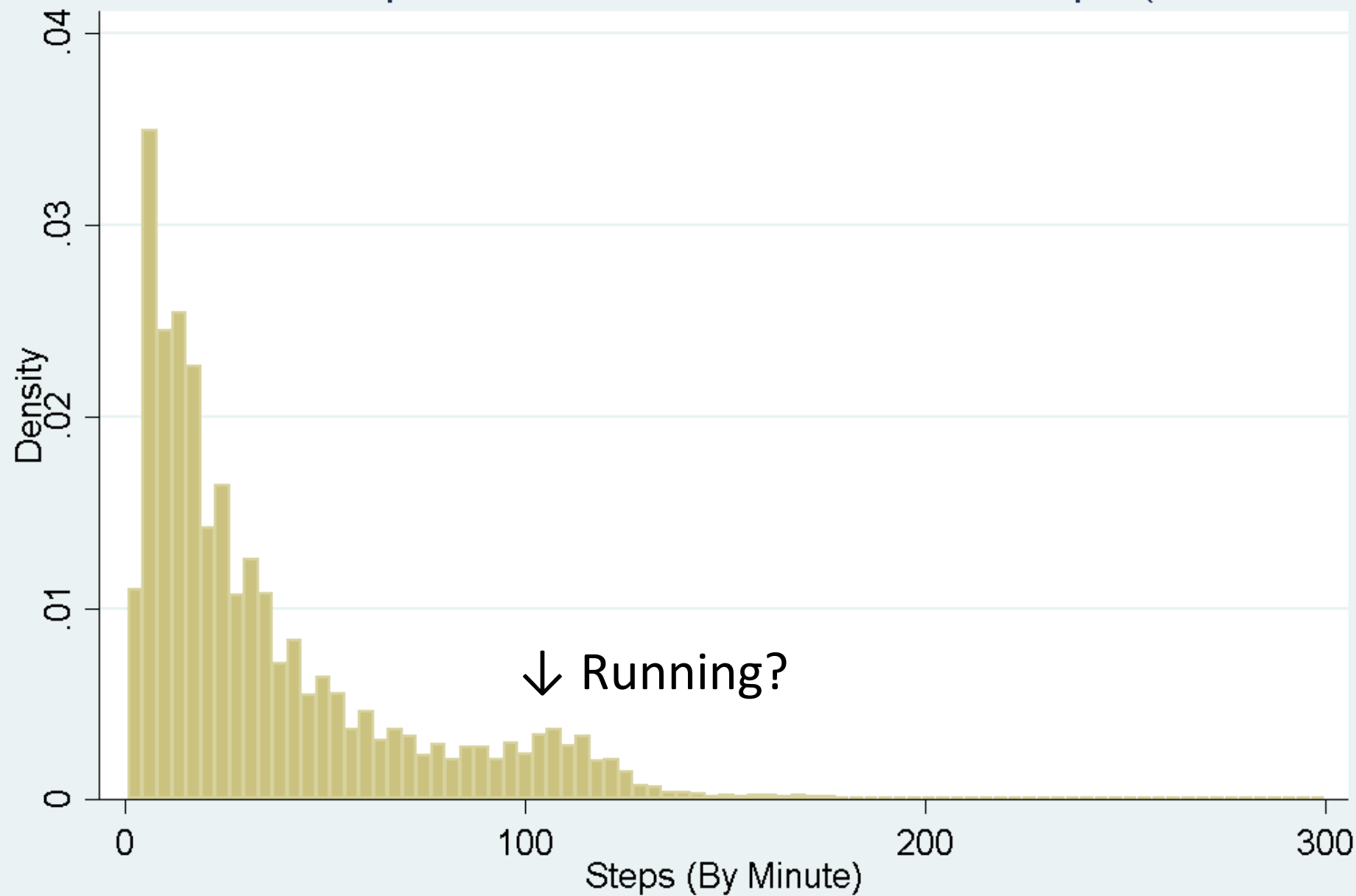
Distribution of minutes in the day with non-zero step counts



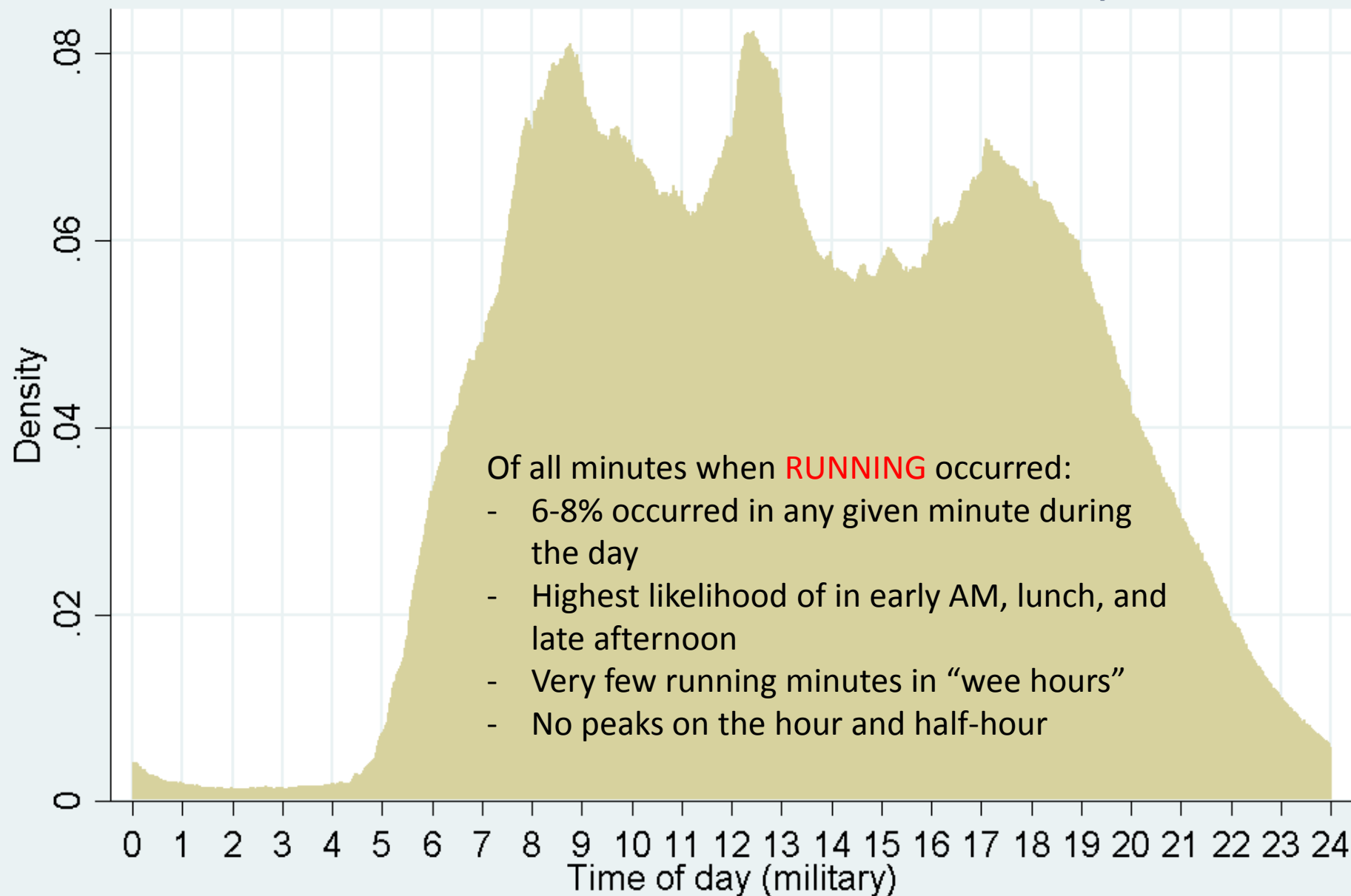
Number of Steps in Minutes with Non-Zero Steps (Trim<300)



Number of Steps in Minutes with Non-Zero Steps (Trim<300)



Distribution of minutes with 100+ steps



Big Data

Analysis commands – how long did it take?

- Simple Stata Commands

```
import delimited "meas_FitBit_Intraday.txt", delimiter("|")
```

```
save "Fitbit_Intraday_imported.dta", replace
```

```
gen newdate = date(date, "DMY")
```

```
gen hours = trunc(minute/60)
```

```
gen minutesleft = minute-(hours*60)
```

```
gen datetime = Cdhms(newdate, hours, minutesleft, 0)
```

```
format datetime %tC
```

```
sort user_id datetime
```

Big Data

Analysis commands – how long did it take?

TIME • Simple Stata Commands

```
1h 33m import delimited "meas_FitBit_Intraday.txt", delimiter("|")

save "Fitbit_Intraday_imported.dta", replace
gen newdate = date(date, "DMY")
gen hours = trunc(minute/60)
gen minutesleft = minute-(hours*60)
gen datetime = Cdhms(newdate, hours, minutesleft, 0)
format datetime %tC

sort user_id datetime
```

Big Data

Analysis commands – how long did it take?

TIME • Simple Stata Commands

```
1h 33m import delimited "meas_FitBit_Intraday.txt", delimiter("|")

12 sec save "Fitbit_Intraday_imported.dta", replace
29 sec gen newdate = date(date, "DMY")
3 sec gen hours = trunc(minute/60)
11 sec gen minutesleft = minute-(hours*60)
7 sec gen datetime = Cdhms(newdate, hours, minutesleft, 0)
<1 sec format datetime %tC

sort user_id datetime
```

Big Data

Analysis commands – how long did it take?

TIME • Simple Stata Commands

```
1h 33m import delimited "meas_FitBit_Intraday.txt", delimiter("|")

12 sec save "Fitbit_Intraday_imported.dta", replace
29 sec gen newdate = date(date, "DMY")
3 sec gen hours = trunc(minute/60)
11 sec gen minutesleft = minute-(hours*60)
7 sec gen datetime = Cdhms(newdate, hours, minutesleft, 0)
<1 sec format datetime %tC

2m 10s sort user_id datetime
```

Big Data

Analysis commands – Which command took longest?

- Simple Stata Commands

```
{ tostring user_id, gen(idstring)  
  tostring newdate, gen(newdatestring)  
  gen id_date = idstring + "_" + newdatestring
```

```
egen totaldaysteps = sum(value), by(id_date)
```

```
keep user_id id_date totaldaysteps
```

```
duplicates drop
```

```
use "Fitbit_Intraday.dta", clear
```

Big Data

Analysis commands – how long did it take?

TIME • Simple Stata Commands

14m 26s {
 tostring user_id, gen(idstring)
 tostring newdate, gen(newdatestring)
 gen id_date = idstring + "_" + newdatestring

12m 16s egen totaldaysteps = sum(value), by(id_date)

8 sec keep user id id_date totaldaysteps

26m 58s duplicates drop

22 sec use "Fitbit_Intraday.dta", clear

Big Data

Analysis commands – how long did it take?

TIME • Simple Stata Commands

3m 15s `sum steps, detail`

2m 46s `hist steps if steps<300`

Big Data

Analysis commands – how long did it take?

- Answer: Annoyingly long, but not impossible...at least not with 230 million rows.

Big Data

- You think that dataset is big...

[Demonstration Video](#)

Big Data

- You think that dataset is big...

[Demonstration Video](#)

- If 5 channels, sampling rate of 10/second:
 - $5 \text{ channels} \times 60 \text{ seconds/minute} \times 10 \text{ samples/second} = 3000 \text{ times the size of the Fitbit Intraday data (per person-day)}$

Other Big Data issues

- Standardization
 - Step counts from Fitbit vs. Jawbone vs. apps
 - Time zones
 - Other “meta-data”
 - Data structure
- Open mHealth provides schema for standardizing
 - <http://www.openmhealth.org/>
 - Ida Sim here at UCSF

Other Big Data issues

```
{
  "systolic_blood_pressure": {
    "value": 160,
    "unit": "mmHg"
  },
  "diastolic_blood_pressure": {
    "value": 60,
    "unit": "mmHg"
  },
  "effective_time_frame": {
    "time_interval": {
      "start_date_time": "2013-02-05T07:25:00Z",
      "end_date_time": "2013-06-05T07:25:00Z"
    }
  },
  "body_posture": "sitting",
  "descriptive_statistic": "maximum"
}
```

- Open mHealth schema for blood pressure measurements (JSON format)

Digital Data – Lessons learned (so far)

- Time series data is the norm
- Data requires cleaning and “wrestling down”
- Even basic interpretation can be a challenge
 - What does a zero or missing value in the time-series sequence mean?
- Data “meaning” often requires meta-data
- Data get to be quite large
 - Inconvenient to analyze using standard statistical packages
 - Inconvenient to move around

Digital Data – Lessons learned (so far)

- Possibly-relevant terms/tools from my perspective as non-data-scientist:
 - “Sharded” databases and distributed analysis?
 - “Redshift”, etc: <https://aws.amazon.com/redshift/>
 - Sequel queries are quicker...
 - “Map/Reduce” – Google analytics?
 - “Hadoop” – for storage, not analytics?
 - “Apache Spark” - <http://spark.apache.org/>

Digital Health – Coming attractions

- Precision Medicine Initiative
- Health ePeople “Eureka” Platform
- MD2K

Precision Medicine Initiative

- <https://www.nih.gov/precision-medicine-initiative-cohort-program>
- “Through advances in research, technology and policies that empower patients, the PMI will enable a new era of medicine in which researchers, providers and patients work together to develop individualized care.”

Precision Medicine Initiative

- Tailoring treatments to individuals
 - Original vision: Genetic testing → precision treatments
 - Revised vision: Includes digital health monitoring
- Large contracts awarded to support mHealth data collection for PMI Cohort
 - We just missed...
- Expect lots of new mHealth data in 5-10 years on large (n=1 million?) cohort

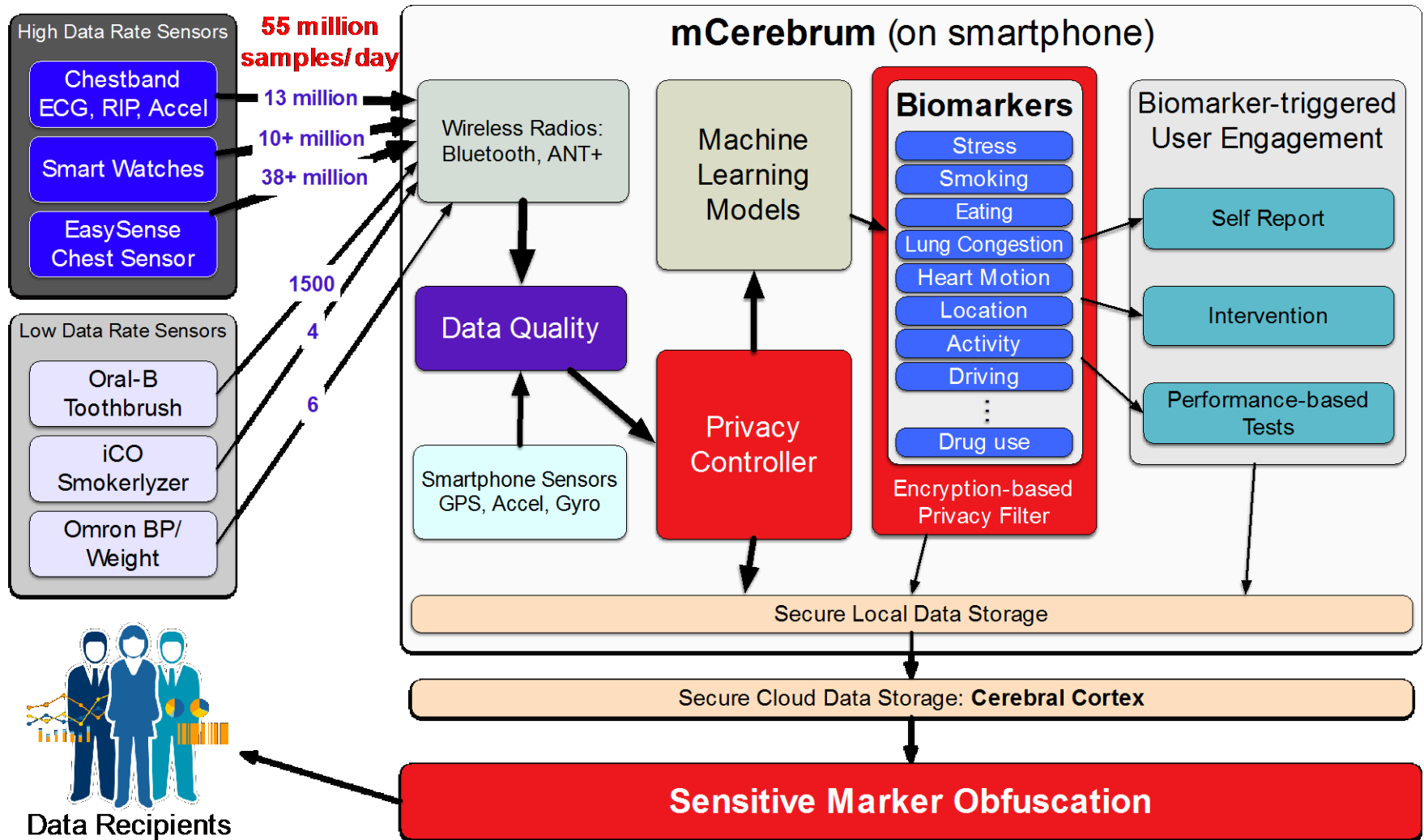
Health ePeople “Eureka” Platform

- Health eHeart → Health ePeople
 - NIH U2C Award to create infrastructure
- Modular/configurable data collection system
 - Surveys, commercial sensors and apps
 - Messaging system for reminders/interventions
 - Branding easily swapped out and modified
 - Scalable back-end on Amazon Web Services
- Cohort/Registry
- De-identified database

MD2K: Mobile Data to Knowledge

- NIH BD2K Program --> MD2K
 - Center of Excellence for Mobile Sensor Data-to-Knowledge (MD2K)
- Develops software, tools, training, and science
 - To gather, analyze, and interpret health-related high-frequency mobile sensor data
 - To develop innovative methods for early detection & prevention of complex chronic diseases,
 - E.g. Sensor-triggered just-in-time intervention for congestive heart failure & smoking cessation

MD2K: Mobile Data to Knowledge



MD2K: Mobile Data to Knowledge

- NIH BD2K Program --> MD2K
 - Center of Excellence for Mobile Sensor Data-to-Knowledge (MD2K)
- Develops software, tools, training, and science
 - To gather, analyze, and interpret health-related high-frequency mobile sensor data
 - To develop innovative methods for early detection & prevention of complex chronic diseases,
 - E.g. Sensor-triggered just-in-time intervention for congestive heart failure & smoking cessation
- “Digital biomarkers”
- On-phone computation

MD2K Team Spread Across 12 Institutions

Computing & Engineering

- Santosh Kumar, *Memphis (PI)*
- Gregory Abowd, Polo Chau, and Jim Rehg, *Georgia Tech*
- Emre Ertin, *OSU*
- Deborah Estrin, *Cornell Tech*
- Tyson Condie, Mani Srivastava, *UCLA*
- Deepak Ganesan, Ben Marlin, *UMass*
- Susan Murphy, *Michigan*

Health Research

- William Abraham, *OSU*
- Mustafa al'Absi, *Minnesota*
- Inbal Nahum-Shani, *Michigan*
- Bonnie Spring, *Northwestern*
- Cho Lam, Dave Wetter, *Rice*
- Vivek Shetty, *UCLA*
- Ida Sim, *UC San Francisco*
- Jaqueline Kerr, *UC San Diego*
- Clay Marsh, *West Virginia*

Memphis-based headquarter hosts a team of 4 software engineers led by Dr. Tim Hnat and a team of 6 staff members led by Joe Biggers.

MD2K Team Spread Across 12 Institutions

Computing & Engineering

- Santosh Kumar, *Memphis (PI)*
- Gregory Abowd, Polo Chau, and Jim Rehg, *Georgia Tech*
- Emre Ertin, *OSU*
- Deborah Estrin, *Cornell Tech*
- Tyson Condie, Mani Srivastava, *UCLA*
- Deepak Ganesan, Ben Marlin, *UMass*
- Susan Murphy, *Michigan*

Health Research

- William Abraham, *OSU*
- Mustafa al'Absi, *Minnesota*
- Inbal Nahum-Shani, *Michigan*
- Bonnie Spring, *Northwestern*
- Cho Lam, Dave Wetter, *Rice*
- Vivek Shetty, *UCLA*
- Ida Sim, *UC San Francisco*
- Jaqueline Kerr, *UC San Diego*
- Clay Marsh, *West Virginia*

Memphis-based headquarter hosts a team of 4 software engineers led by Dr. Tim Hnat and a team of 6 staff members led by Joe Biggers.

This is the guy in the video!

Take Home Points

- Digital Health is here to stay
- It will take a while to make it useful
 - Think of lag between whole genome sequencing to clinical utility and health
- Not everyone uses digital technologies
 - Could we be increasing disparities?
- Tech is moving target
- New tools/infrastructure/data coming soon
- Most skills we teach is relevant, but bigness of data may require different skills...
- Start planning your own digital health study!?

Take Home Points Questions?

- Digital Health is here to stay
- It will take a while to make it useful
 - Think of lag between whole genome sequencing to clinical utility and health
- Not everyone uses digital technologies
 - Could we be increasing disparities?
- Tech is moving target
- New tools/infrastructure/data coming soon
- Most skills we teach is relevant, but bigness of data may require different skills...
- Start planning your own digital health study!?