Summer Course Survey Research: Advanced Survey Design

Bella Struminskaya & Peter Lugtig

Department of Methodology & Statistics, Utrecht University

Slides by Struminskaya & Keusch

Utrecht University

Data collection using apps and sensors

Bella Struminskaya & Peter Lugtig

Department of Methodology & Statistics, Utrecht University

Exercise

Why would you add smartphone sensors, apps, and wearables to surveys?

Why use sensors for data collection

Potential benefits of apps, sensors, & wearables

1. Taking advantage of technology that is widely used in society
    - High smartphone penetration & quantified-self movement
    - Device present in same physical and social context as user
    - Moving from small scale lab studies to larger scale field studies

Potential benefits of apps, sensors, & wearables

1. Taking advantage of technology that is widely used in society
2. Multiple (new) forms of measurement on a single device
   - *In situ* measurement (e.g., EMA/ESM)
   - *Passive measurement* with sensors (e.g., automatic collection of location and activity)
   - Use of other device features for *active measurement* (e.g., photos, videos)
   - Smartphone as *hub* for other devices (e.g., smart watch, smart scale, via Bluetooth)

Potential benefits of apps, sensors, & wearables

1. Taking advantage of technology that is widely used in society
2. Multiple (new) forms of measurement on a single device
3. More detailed data (frequency and intensity)
    - High frequency of measurement (e.g., intensive longitudinal measurement, passive measurement)
    - Much more fine-grained data than in traditional longitudinal designs
    - New types of information that cannot be self-reported (e.g., different stages of sleep)

Potential benefits of apps, sensors, & wearables

1. Taking advantage of technology that is widely used in society
2. Multiple (new) forms of measurement on a single device
3. More detailed data (frequency and intensity)
4. Unobtrusive, direct measurement should lead to more accurate estimates
    - Less self-report = Less recall error
    - Less self-report = (Potentially) less social desirability
    - Less self-report = Less data entry error

Potential benefits of apps, sensors, & wearables

1. Taking advantage of technology that is widely used in society
2. Multiple (new) forms of measurement on a single device
3. More detailed data (frequency and intensity)
4. Unobtrusive, direct measurement should lead to more accurate estimates
5. Less response burden
   - Fewer survey questions have to be asked about (Harari et al. 2017)...
     - Smartphone-mediated behaviors (e.g., # of calls & text messages, Internet browsing, app use)
     - Non-mediated behaviors (e.g., physical activity, sleep, movement, travel)
     - Daily activities (e.g., food intake, expenditure)
   - But what about other burden? - Consent, compliance, privacy, etc.

Potential benefits of apps, sensors, & wearables

1. Taking advantage of technology that is widely used in society
2. Multiple (new) forms of measurement on a single device
3. More detailed data (frequency and intensity)
4. Unobtrusive, direct measurement should lead to more accurate estimates
5. Less response burden
6. Collecting data at scale
    - ~22,000 volunteer iPhone users downloaded *Mappiness* app and shared activities and affect (EMAs) plus geolocation (GPS) for 6 months (MacKerron & Mourato 2013)
    - 650 members of existing longitudinal study downloaded *IAB-SMART* app and responded to mini-surveys plus shared location, physical activity, and smartphone use data for 6 months (Kreuter et al. 2020)
    - >100,000 participants of the UK Biobank study wore wrist accelerometer for 7 days (Doherty et al. 2017)

Potential benefits of apps, sensors, & wearables

1. Taking advantage of technology that is widely used in society
2. Multiple (new) forms of measurement on a single device
3. More detailed data (frequency and intensity)
4. Unobtrusive, direct measurement should lead to more accurate estimates
5. Less response burden
6. Collecting data at scale
7. New research questions?

Potential challenges of apps, sensors, & wearables

1. Coverage
    - “Ubiquity Myth” (Couper 2019)
    - Age, education, gender…
    - “2nd-level digital divide”

Potential challenges of apps, sensors, & wearables

1. Coverage
2. Nonparticipation
   - Willingness
   - Ability
   - Adherence to study protocols

Potential challenges of apps, sensors, & wearables

1. Coverage
2. Nonparticipation
3. Privacy & ethics
    - What concerns do people have?
    - “Privacy paradox”

Potential challenges of apps, sensors, & wearables

1. Coverage
2. Nonparticipation
3. Privacy & ethics
4. Measurement
   - Data not free of error
   - Technical issues and human behavior can lead to missings and implausible readings

What can we measure and what types of research questions can we answer

Different devices with sensors

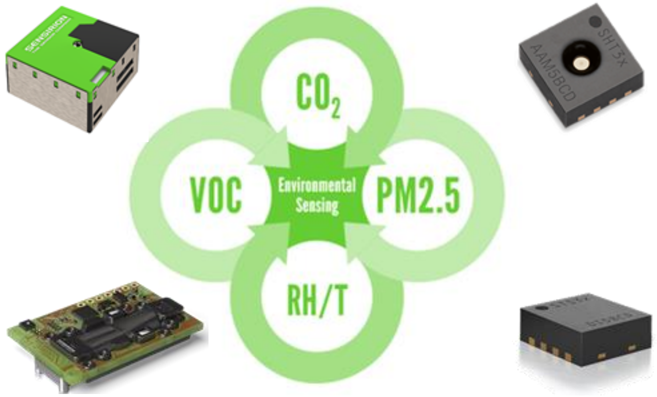
Source: https://www.youtube.com/watch?v=FEr9D2glDXA

Source: https://www.techradar.com/news/wearables/10-best-fitness-trackers-1277905

Sources: http://www.canadagps.com/CanmoreGT-750FL_Sirf4.html
https://www.laserinst.com/trimble-geo7x-handheld/

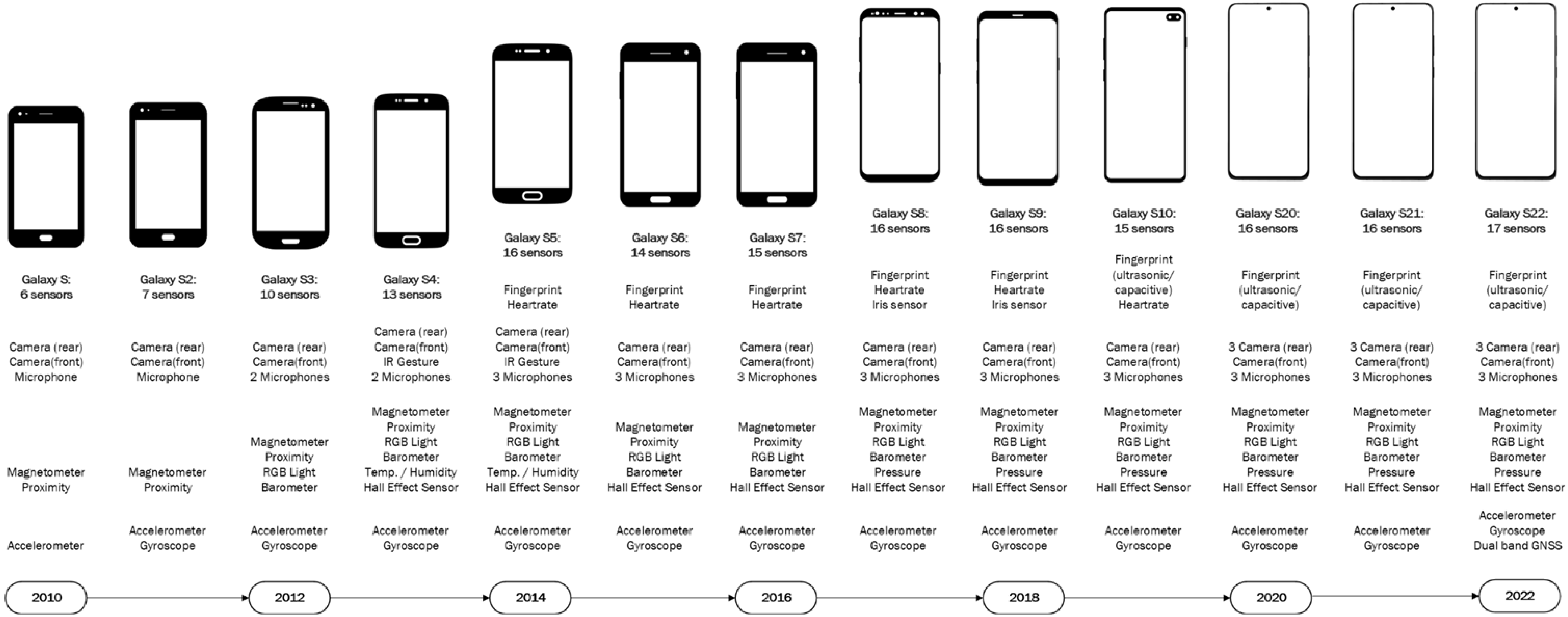
Source: https://www.techradar.com/news/wearables/best-smart-watches-what-s-the-best-wearable-tech-for-you-1154074

Sources: https://www.actigraphcorp.com/actigraph-wgt3x-bt/, https://www.activinsights.com/products/geneactiv/



Source: https://www.sensirion.com/en/environmental-sensors/

Smartphones & sensors



Struminskaya & Keusch. (2023)

Native smartphone sensors

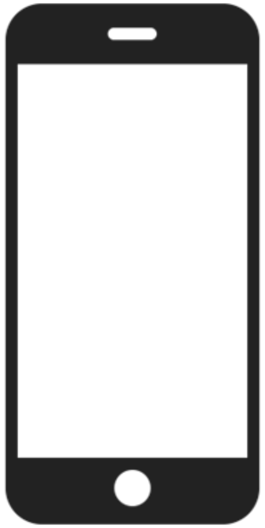
NFC

Heart rate sensor

Bluetooth

Air humidity sensor

Thermometer



Proximity sensor

Wi-Fi

Microphone

GPS

Light sensor

Cellular network

Fingerprint sensor

Camera

Barometer

Compass

Accelerometer

Pedometer

Gyroscope

Exercise

What can you measure with these sensors?

Native smartphone sensors

Proximity

NFC

Bluetooth

Heart rate sensor

Air humidity sensor

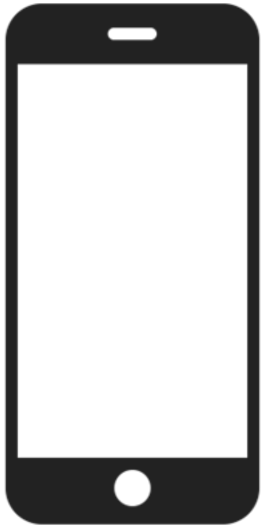
Thermometer

Location

Wi-Fi

GPS

Cellular network



Proximity sensor

Microphone

Light sensor

Ambience/ Environment

Camera

Fingerprint sensor

Barometer

Compass

Accelerometer

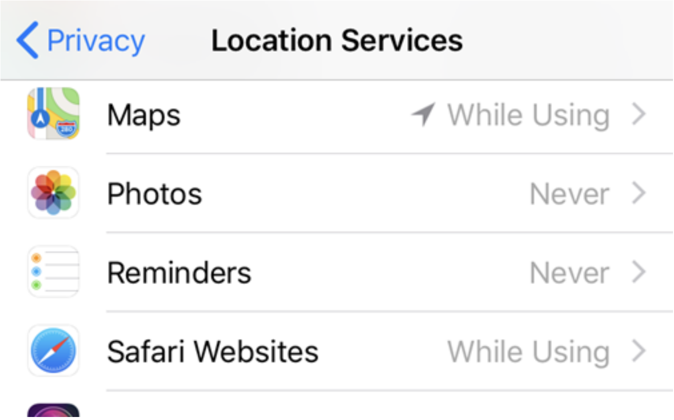
Pedometer

Gyroscope

Physical activity

In-browser (Java/HTML) vs. App measurement

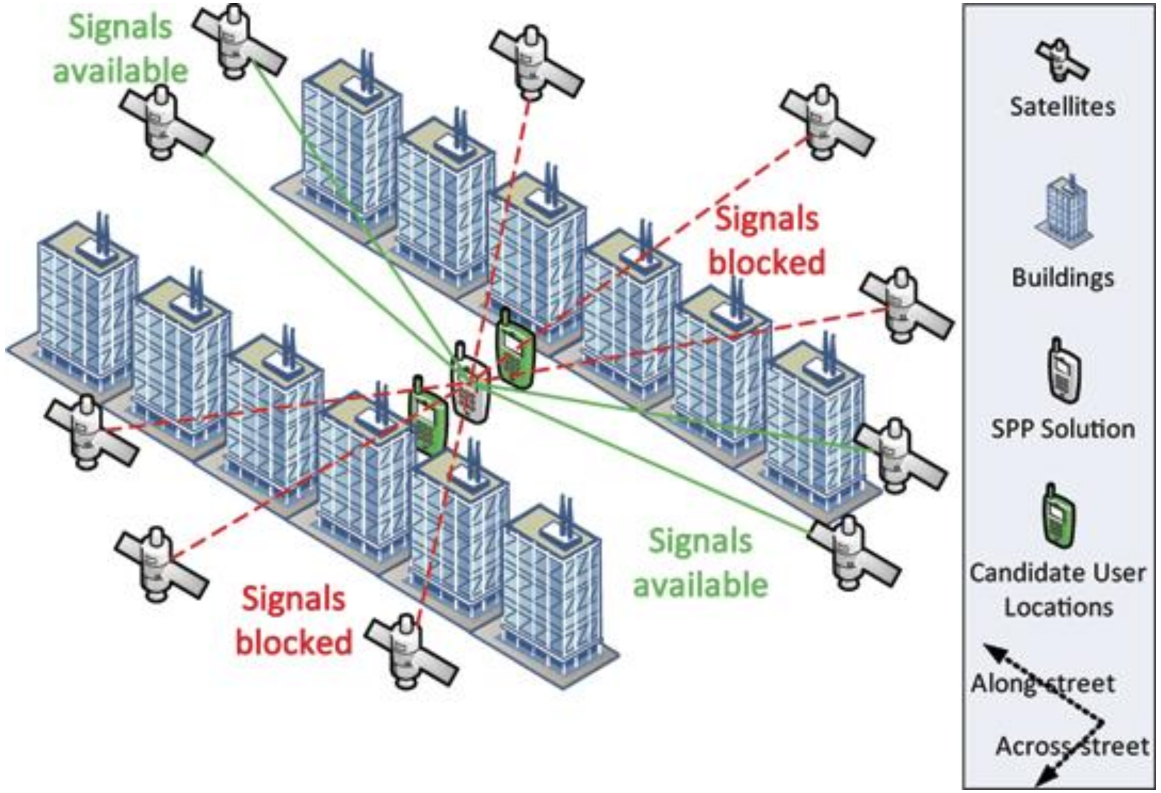
- Data collection implemented through specific website
- Sensing limited to when website is open
- No need to install app but participant needs to consent and change permission in settings



- Data collection “in background”
  - Amount and frequency of data collection depends on OS and version
- Participant needs to download app and give permissions
  - For each sensor individually
- “Offline” data collection
- Triggering of questions (EMA, geo-fencing)
- Battery life & other technical issues

Geolocation

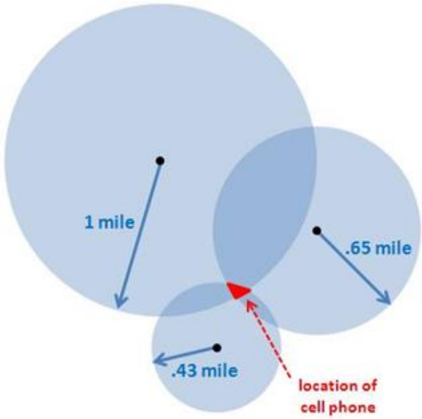
- GPS
  - Provides coordinates in longitude & Latitude
  - Based on distance (= rate x time) to at least 4 satellites
  - Newest generation has accuracy within 30 centimeters
  - Works without cell/Internet connection
  - Performs worse in 'urban canyons', indoors, & underground
  - Constant tracking is very battery-draining



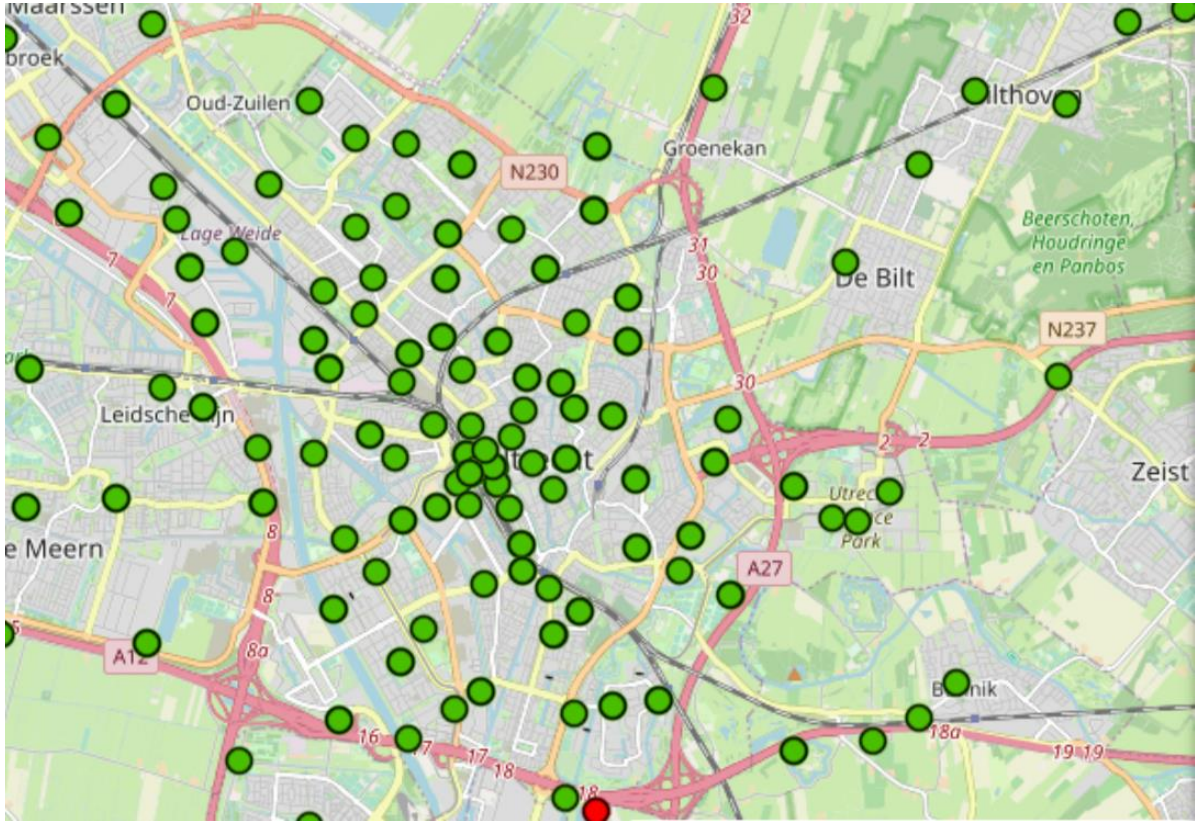
Source: https://www.gpsworld.com/wirelesspersonal-navigationshadow-matching-12550/

Geolocation

- GPS
- Cellular network
  - Multilateration of radio signals between (several) cell towers
  - Works even if GPS is turned off
  - If there is no signal then location information will be missing



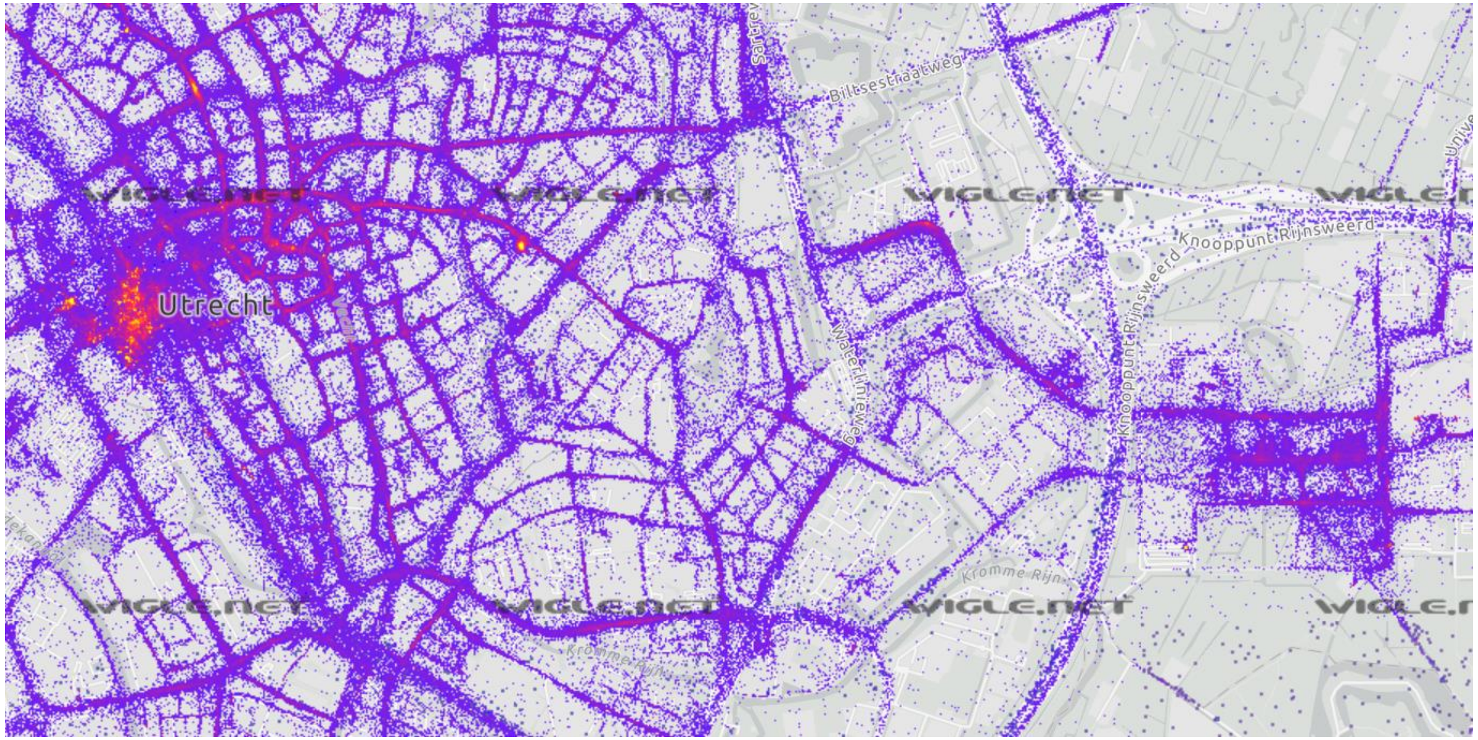
Source: https://searchengineland.com/cell-phone-triangulation-accuracy-is-all-over-the-map-14790



Source: https://www.cellmapper.net

Geolocation

- GPS
- Cellular network
- Wi-Fi
  - Inferring location from Wi-Fi access points (AP)
  - Can overcome problem of ‘urban canyons’ and indoor tracing

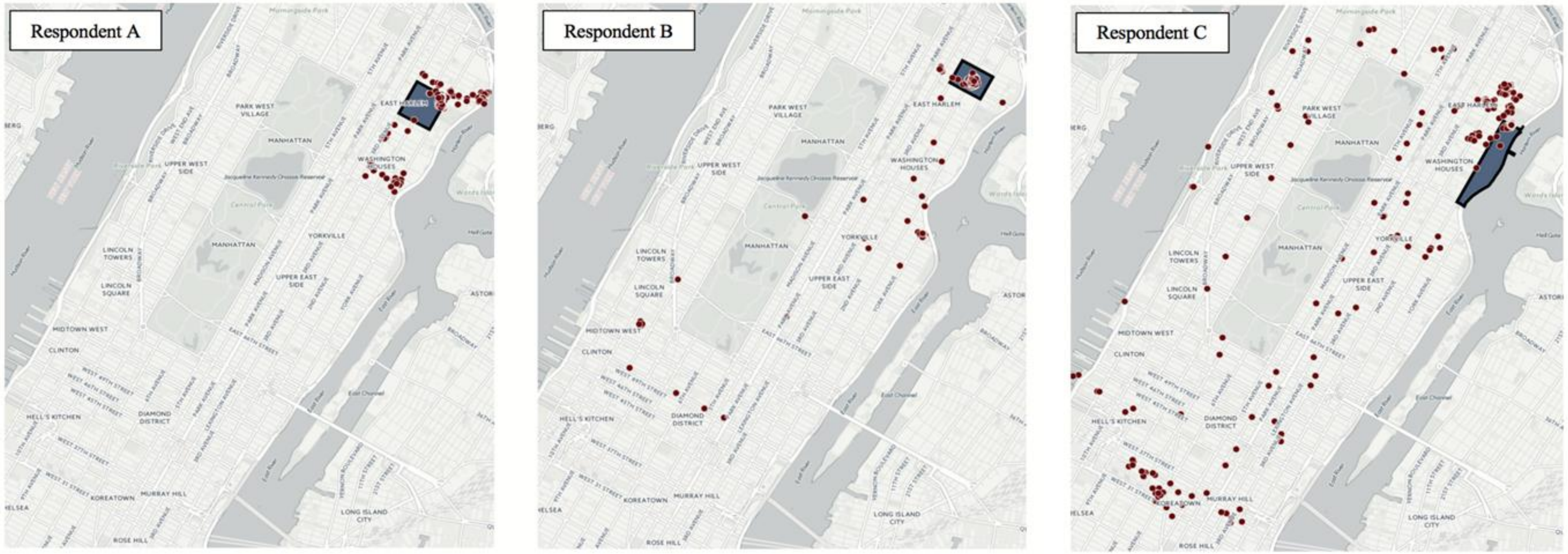


Source: https://www.wigle.net

Example: Aging in activity space

(York Cornwell & Cagney 2017, 2020)

- *Real-time Neighborhoods and Social Life Study* (RNSL)
- 60 participants aged 55+ in NYC provided with iPhones to carry for 7 days
- GPS-tracking (every 5 min) from 9 a.m. to 9 p.m and four EMAs per day



Example: Aging in activity space

(York Cornwell & Cagney 2017, 2020)

- Activity spaces vary considerably in size
- Participants spent ~40% of their time outside their residential tracts
  - On average >10 min in 9+ tracts
- Activity spaces larger among younger and more advantaged social groups (i.e., whites, those with college degree, car owners)
- Participants with less education and lower incomes spend more time outside
  of their residential tracts
- Four main activities outside of residential tracts
  - Shopping, exercising, socializing, participating in social groups or activities
- Poverty rates in nonresidential tracts lower than in residential tracts
- Higher concentrated disadvantage in an area associated with higher odds of self-reporting pain

Example: How do people find work after prison?

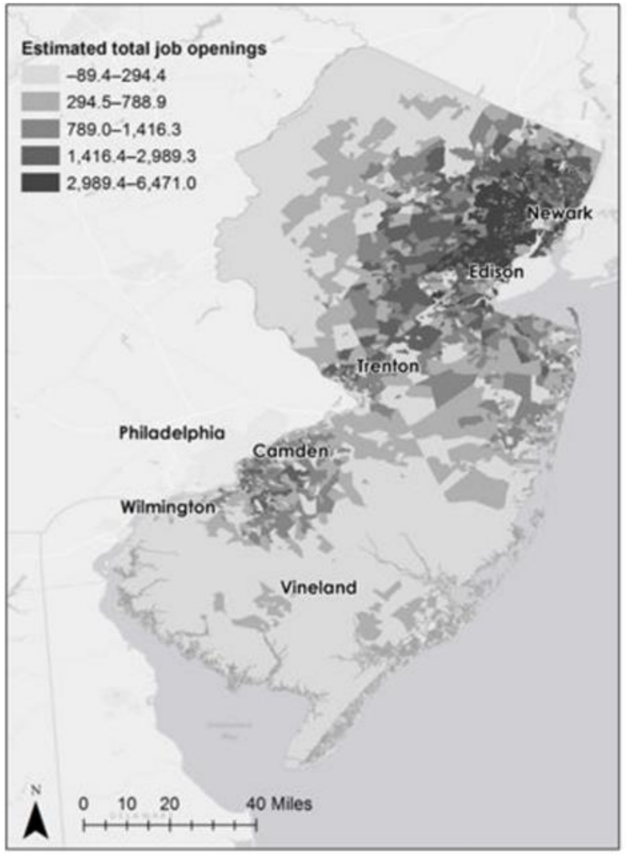
(Sugie 2018; Sugie and Lens 2017)

- Newark Smartphone Reentry Project (NSRP) 2012-2013
  - N = 133 with 8,000 daily observations (89% response, 1.5% noncompliance)
  - 3 months of data collection
- Men recently released from prison
  - Difficult group to follow due to unstable circumstances
- Loaner smartphones (Android)
- Surveys twice a day (EMA) about social interaction, job search & work, and emotional well-being
- Sensing
  - GPS location
  - Calls and messaging (encrypted)
- Survey triggered by calls/messages from new telephone numbers

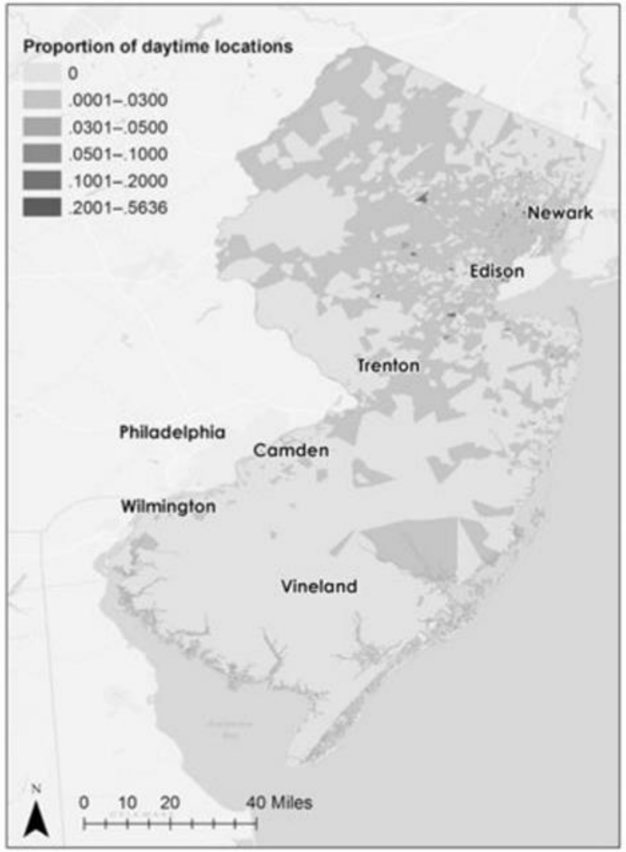
Example: How do people find work after prison?

(Sugie 2018; Sugie and Lens 2017)

- Spatial mismatch
  - Low-skilled, nonwhite job seekers within central cities, job opportunities in outlying areas
- Hypothesis
  - Parolees lack info on job openings, are geographically restricted, unable to travel to find work
- Findings
  - Residential mismatch lengthens time to employment
  - But mobility can compensate for residential deficits



Job openings

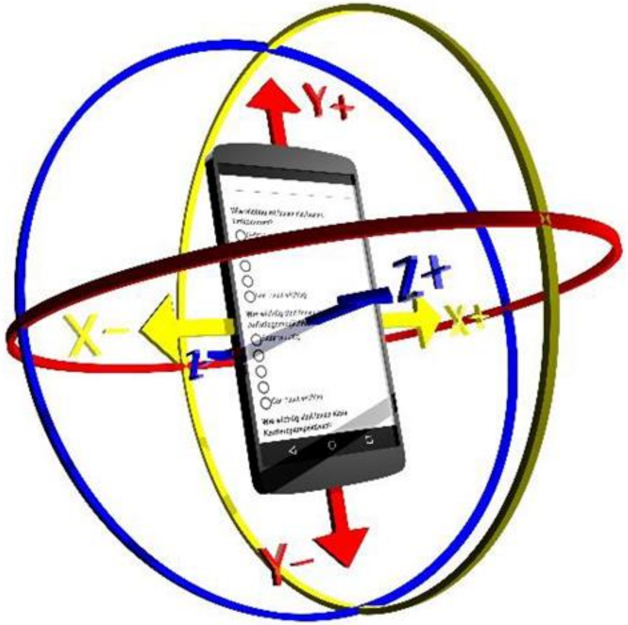


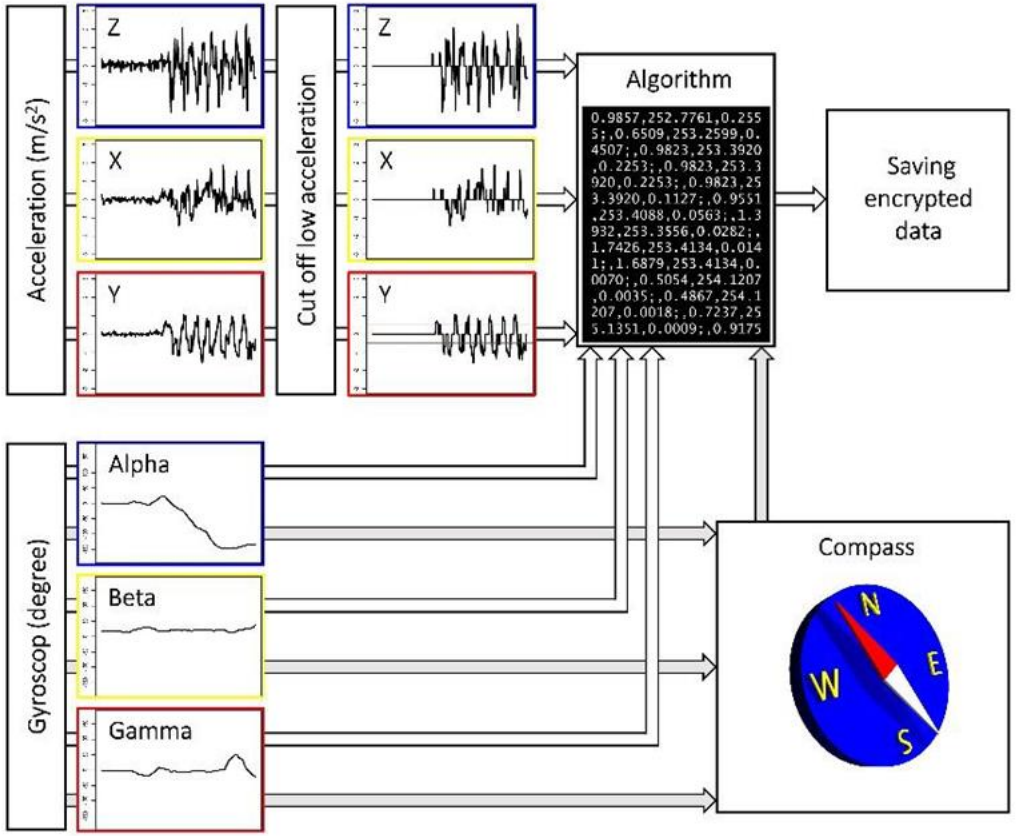
Daytime locations of parolees

Physical activity

- Accelerometer
- Gyroscope

Source: https://www.techradar.com/news/wearables/10-best-fitness-trackers-1277905





Schlosser et al. (2019)

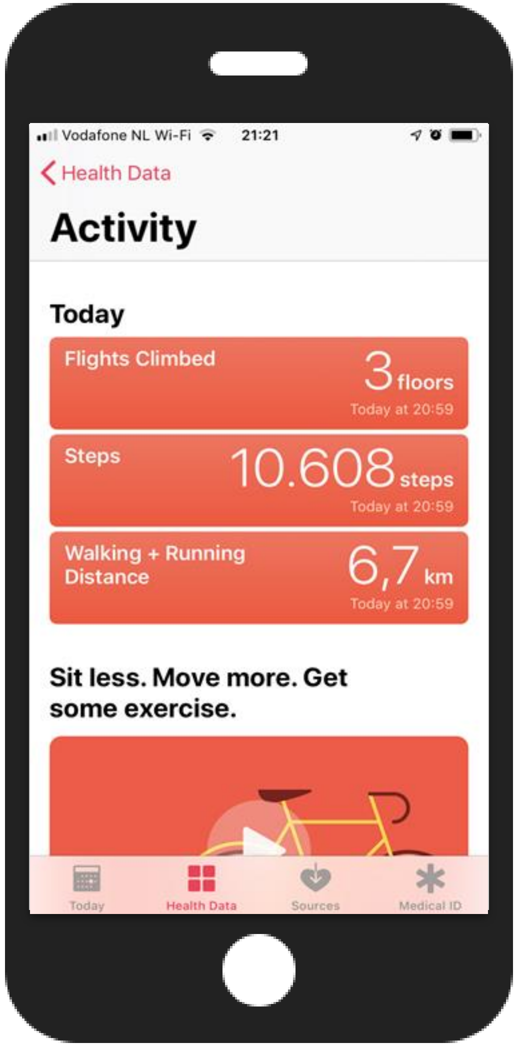
Sources: https://www.actigraphcorp.com/actigraph-wgt3x-bt/, https://www.activinsights.com/products/geneactiv/

Physical activity

- Accelerometer
- Gyroscope

and

- Magnetometer
  - Serves as compass
- Barometer
  - Allows to track changes in elevation



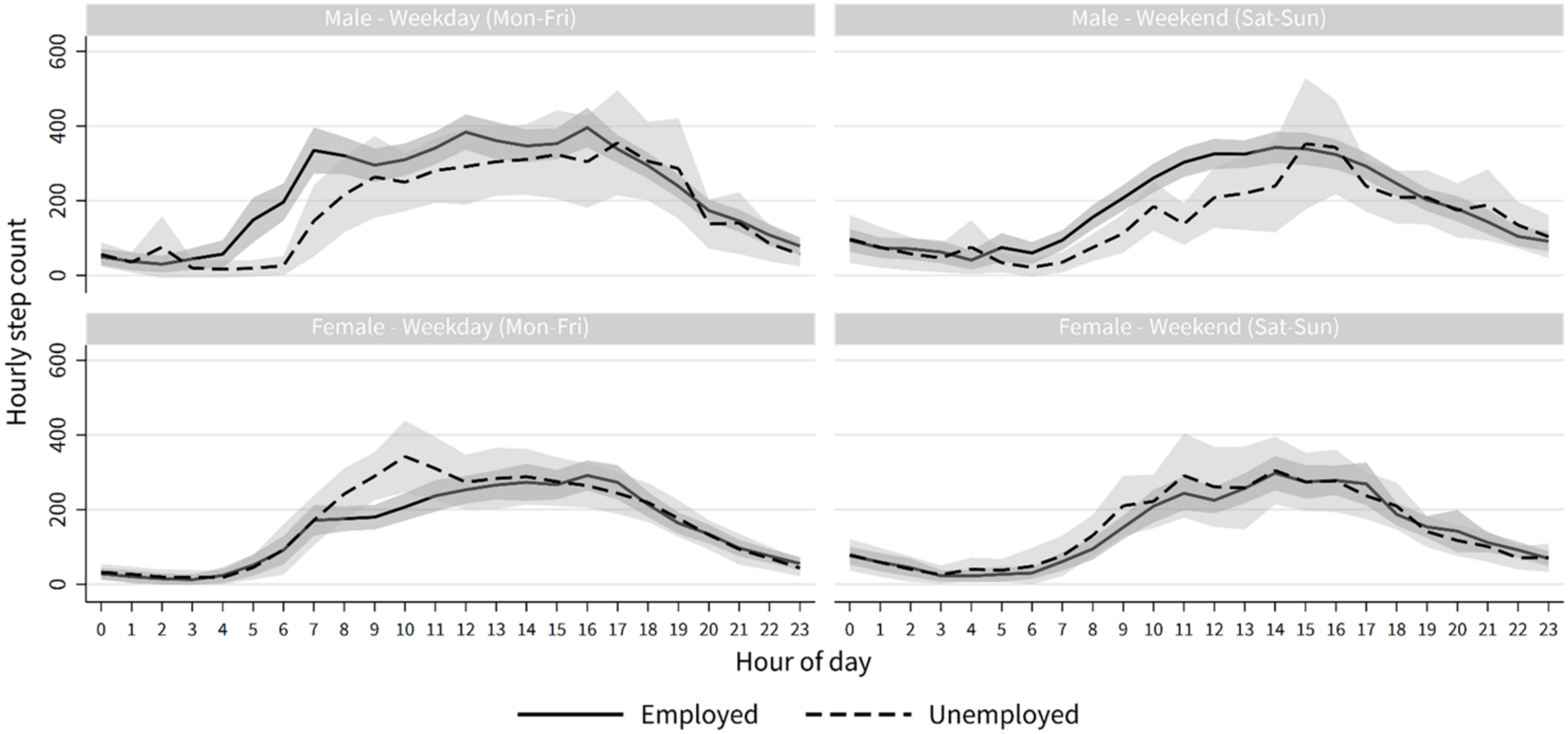
Example: What are the effects of unemployment?

(Kreuter et al. 2020)

- ~650 Android smartphone owners from German panel study “Labour Market and Social Security” (PASS) downloaded *IAB-SMART* app for 6 months
- Survey questions triggered by…
  - Schedule: *Q*s about affective impact of daily smartphone use, Big 5 personality, employment and job search activities, use of smartphones in everyday life, etc.
  - Geolocation: 400 job centers - *Q*s about visit to job center
- Five passive data collection modules:
  - Location using GPS, Wi-Fi, and cellular sensors every 30 min
  - Activity and means of transportation (e.g., walking, biking, riding in/on a motorized vehicle) using accelerometer and pedometer every 2 min
  - Call and texting behavior using phone and SMS logs
  - Use of apps installed on smartphone
  - Social network characteristics from contact lists

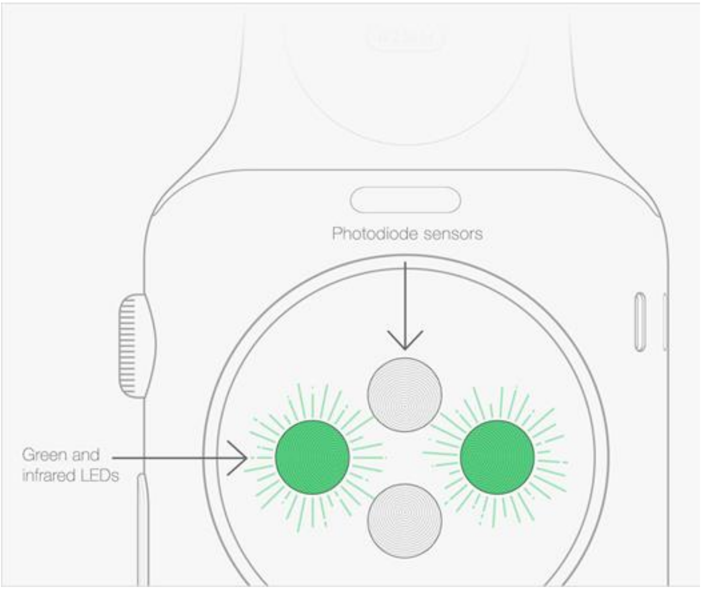
Example: What are the effects of unemployment?

(Bähr et al. in preparation)

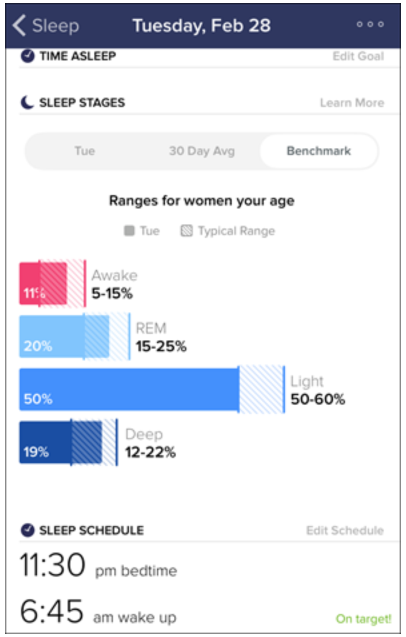


Heart-rate

- Most wristbands use LED-based system
  - Light "shines" onto skin, sensor detects blood volume changes
  - "... finely-tuned algorithms are applied to measure heart rate automatically and continuously..." (https://help.fitbit.com/articles/en_US/Help_article/1565)
  - Samsung Galaxy S uses similar system
- Used in combination with accelerometer to determine sleep phases (e.g., on Fitbit)



Source: https://exist.io/blog/fitness-trackers-heart-rate/



Source: https://help.fitbit.com/articles/en_US/Help_article/2163

Sound & light

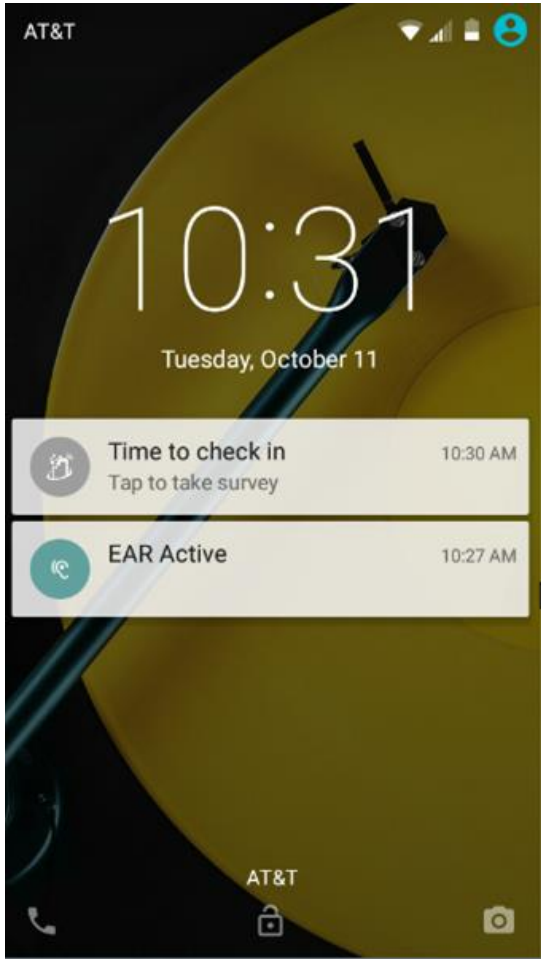
- Microphone
  - “Actively” records answers to survey questions
  - “Passively” measures ambient noise (e.g., clutter), music, and conversations
  - To preserve privacy, classifiers determine that participant is, for example, “around conversation” but not able to reconstruct content or to identify individual speakers
- Light sensor
  - Used to adjust display brightness
  - In combination with other sensors (e.g., accelerometer, microphone) infers idle state of phone/user & sleep

Source: https://www.theverge.com/circuitbreaker/2017/9/15/16307802/apple-iphone-x-features-specs-best-worst

Example: *Daily Experiences and Well-being Study*

(Fingerman et al. 2020)

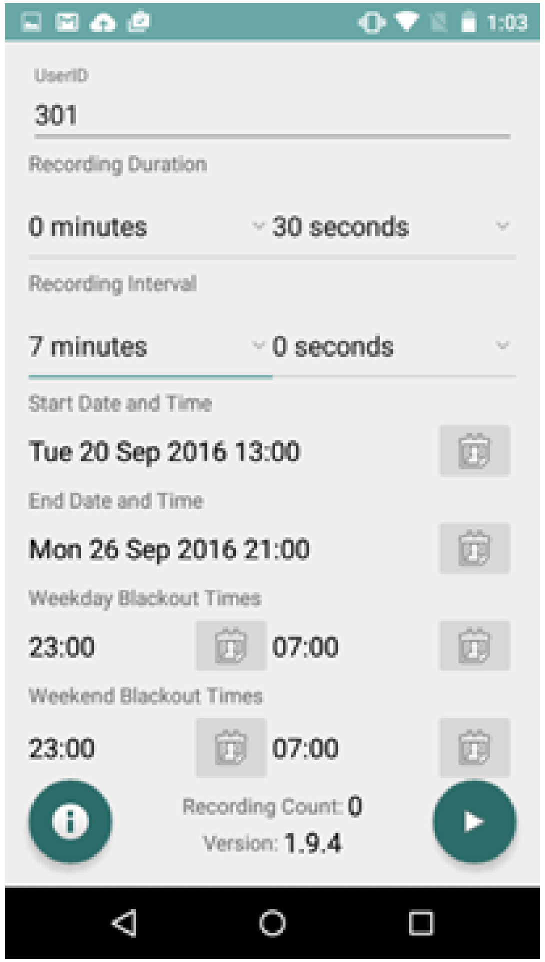
- Telephone screening to identify home-dwelling aged 65+ in Austin, TX (n=333)
  - Oversample of Blacks and Hispanics
  - Without cognitive impairment, not working full-time
- Goal: Study influence of social engagement on physical activity, health, and cognitive status
- In-home interview followed by 5 days of:
  - Actigraphy
  - Loaner Android device with apps to <u>record sound</u> and prompt for ecological momentary assessment (EMA) - no other smartphone functionality
- Daily reminder phone calls & in-home assistance



Example: Electronically Activated Recorder (EAR)

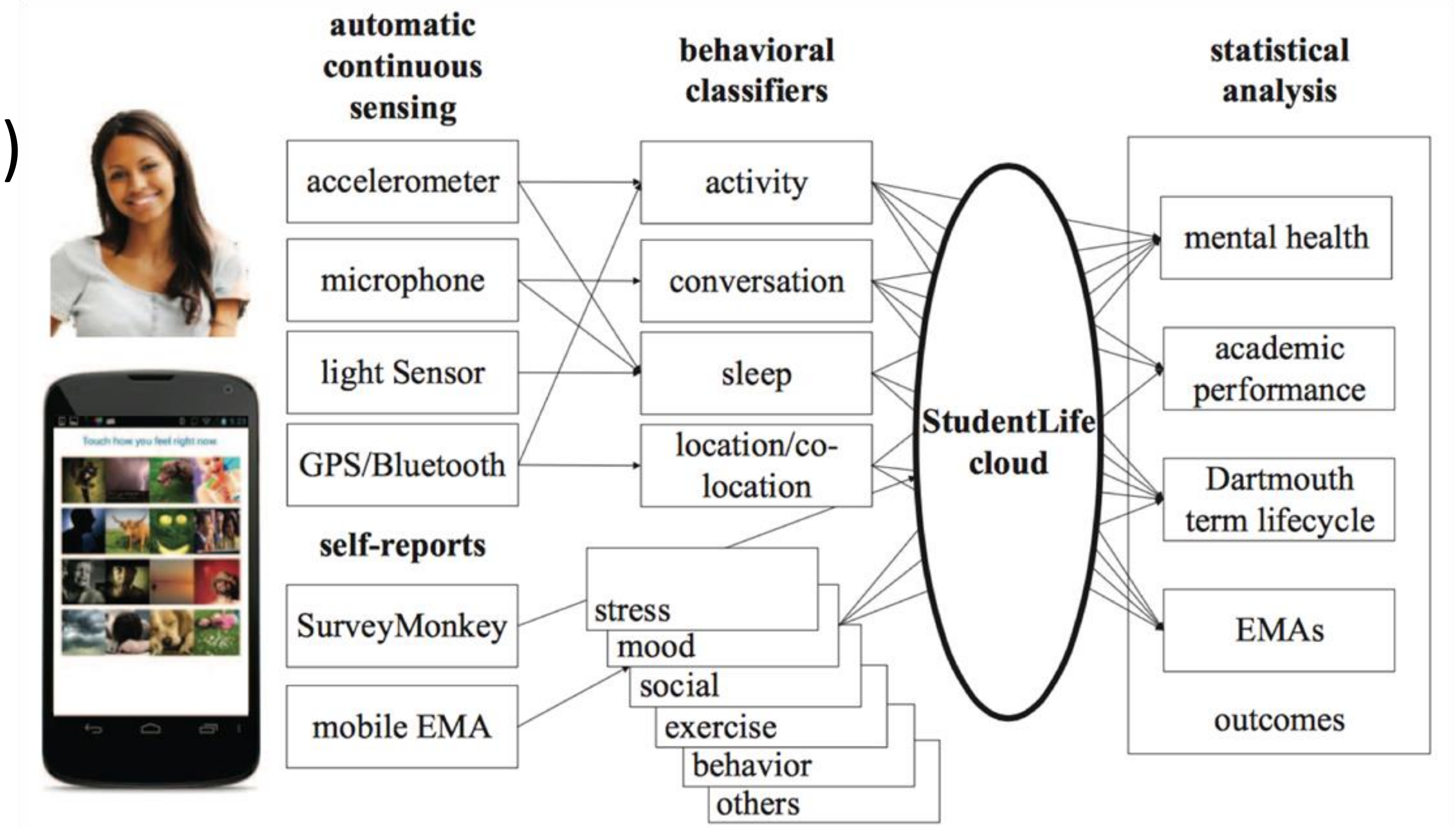
(Fingerman et al. 2022)

- During in-home interview, interviewers entered settings in *EAR* app on phone
  - 30s of recordings every 7 min during waking hours
  - Total of 135,078 audio files
- Devices obtained by interviewer on day 5
- Interviewers responsible for upload and transfer of all data from various devices
- Coders rated each file containing sound for presence of television
- Findings:
  - More TV watching when alone
  - More loneliness reported during periods of TV watching



Does mental health of students change over the course of a term? (Wang et al. 2014)

- 48 students (U.S. college)
- 10 weeks
- Android phones
  (37 provided, 11 own)
- EMA 8 times a day
- Pre- and post-survey



Does mental health of students change over the course of a term? (Wang et al. 2014)

- Students who sleep less, interact less with other students, have fewer co-locations with others more likely to be depressed
- Students around more conversation and students who move around less while on campus do better academically

Correlation with depression

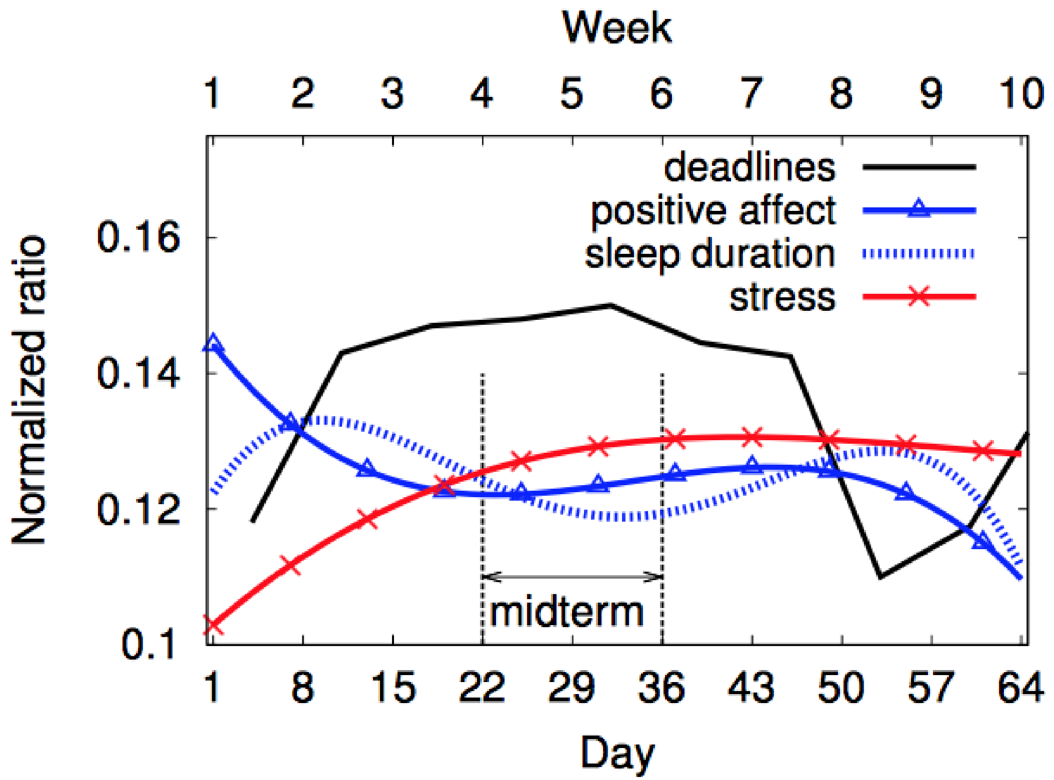
automatic sensing data	r	p-value
sleep duration (pre)	-0.360	0.025
sleep duration (post)	-0.382	0.020
conversation frequency during day (pre)	-0.403	0.010
conversation frequency during day (post)	-0.387	0.016
conversation frequency during evening (post)	-0.345	0.034
conversation duration during day (post)	-0.328	0.044
number of co-locations (post)	-0.362	0.025

Correlation with academic performance

academic performance	Sensing Data	r	p-value
spring GPA	conversation duration (day)	0.356	0.033
spring GPA	conversation frequency (day)	0.334	0.046
spring GPA	indoor mobility	-0.361	0.031
spring GPA	indoor mobility during (day)	-0.352	0.036
spring GPA	indoor mobility during (night)	-0.359	0.032
overall GPA	activity duration	-0.360	0.030
overall GPA	activity duration std deviation	-0.479	0.004
overall GPA	indoor mobility	-0.413	0.014
overall GPA	indoor mobility during (day)	-0.376	0.026
overall GPA	indoor mobility during (night)	-0.508	0.002
overall GPA	number of co-locations	0.447	0.013

Does mental health of students change over the course of a term? (Wang et al. 2014)

- Start of term: high positive affect and conversation levels, low health, healthy sleep, and daily activity
  patterns
- As term progresses: stress rises; activity, sleep, conversation, and positive affect, visits to the gym and attendance drop



(a) EMA and sleep data

Proximity - Bluetooth

- Short-range communication between devices up to 30
  - e.g., hands-free devices, audio speakers, printers
- Enabled healthcare devices can connect to smartphones or other hubs to transmit data
  - e.g., weight, blood pressure, temperature, heart rate, etc.
- Beacons = small Bluetooth transmitters
  - Need to be dispatched by researcher
  - Bluetooth needs to be activated on receiving device
  - Great for indoor tracking

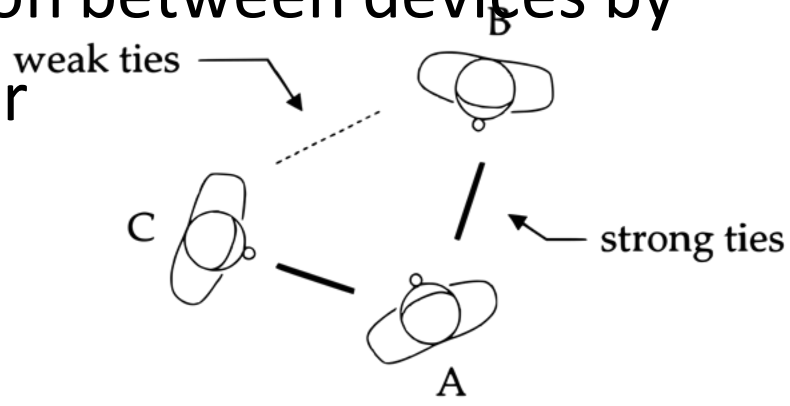
Source: https://www.renesas.com/jp/en/solutions/proposal/bluetooth-low-energy.html

Source: Silvana Jud

Source
http://www.fenc.com/dynafeed/

Proximity - RFID & NFC

- Radio-frequency identification (RFID): electromagnetic fields to automatically identify and track tags attached to objects ~1 meter (3 feet)
    - e.g., assembly lines, merchandise in warehouses, livestock
- Near-field communication (NFC): communication between devices by bringing them within 4 cm (1.6 in) of each other
    - More secure than RFID
    - e.g., contactless payment, data transfer, key cards
- All of them (incl. Bluetooth) can be used to track “social ties”



Source: https://upload.wikimedia.org/wikipedia/commons/2/2a/Weak-strong-ties.svg

Example: How do people interact in large social networks? (Stopczynski et al. 2014)

- *Copenhagen Networks Study*: 1,000 smartphones handed out to Danish students
- Extensive questionnaire upon enrollment: 310 questions on topics from public health, psychology, anthropology, and economics
- Combination of Bluetooth and Wi-Fi networks to collect information about absolute location and relative location to each other
    - Additional data sources: call and text logs, social media data

Example: How do people interact in large social networks? (Stopczynski et al. 2014)

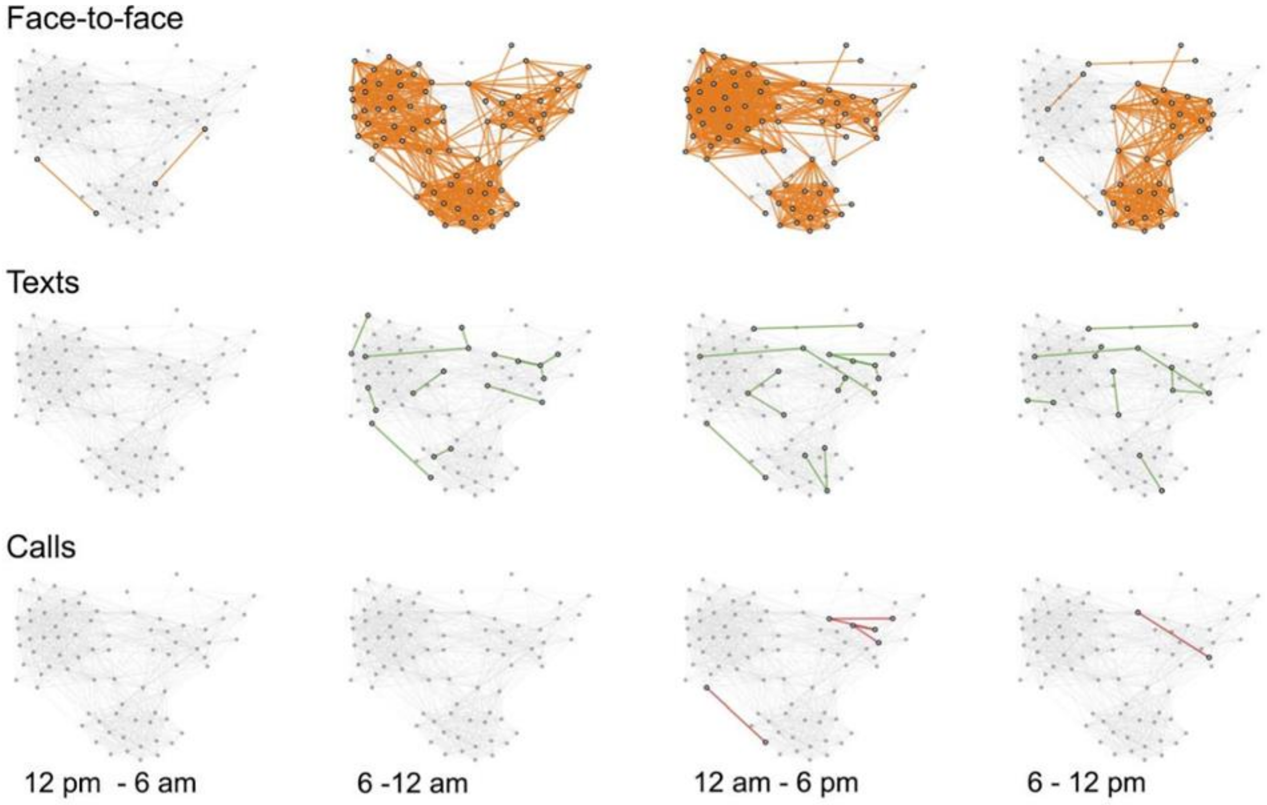
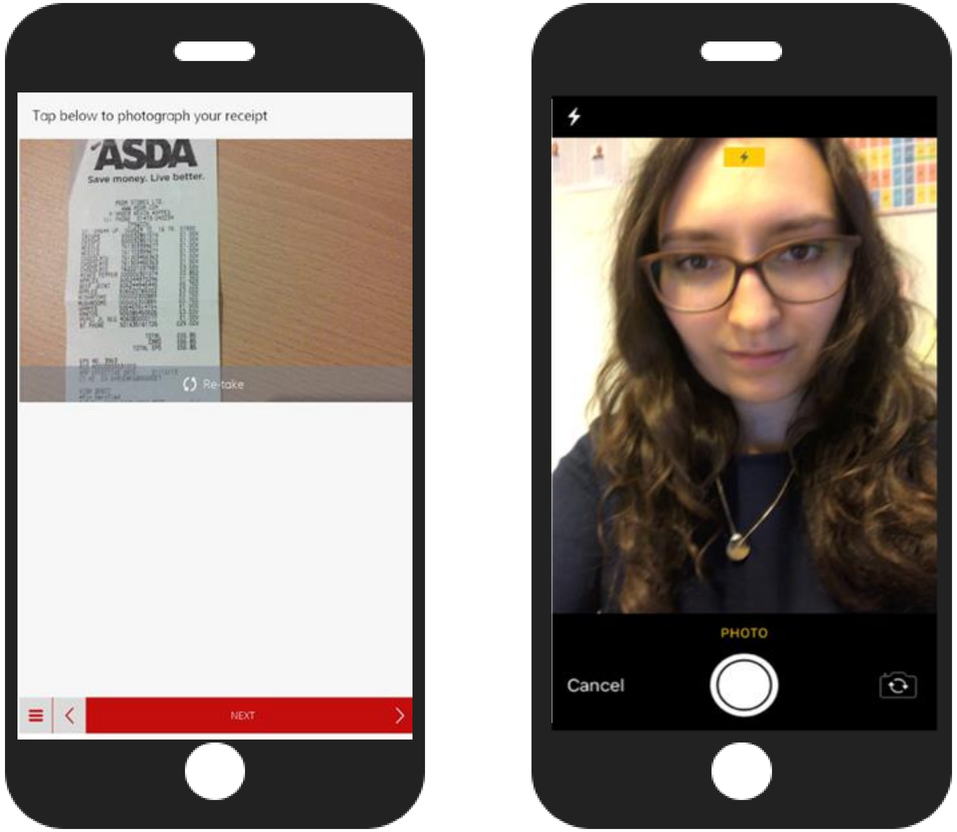


Figure 11. Daily activations in three networks. One day (Friday) in a network showing how different views are produced by observing different channels.

Images

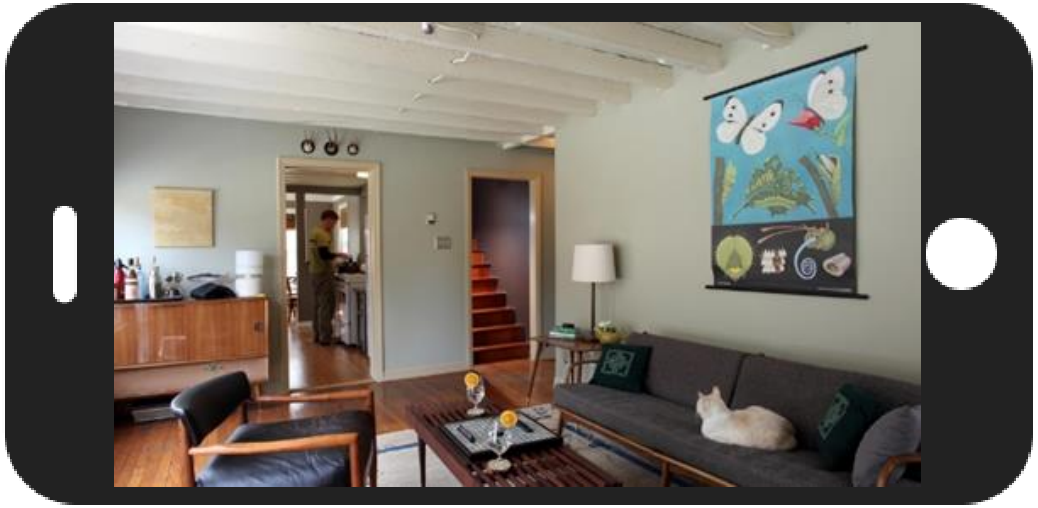
- Photos
  - Food, receipts, physical surroundings, etc.
- Video
- Barcodes
- Linear distance (iPhone Measure app)



Jäckle et al. (2018)

Example: Taking pictures of surrounding

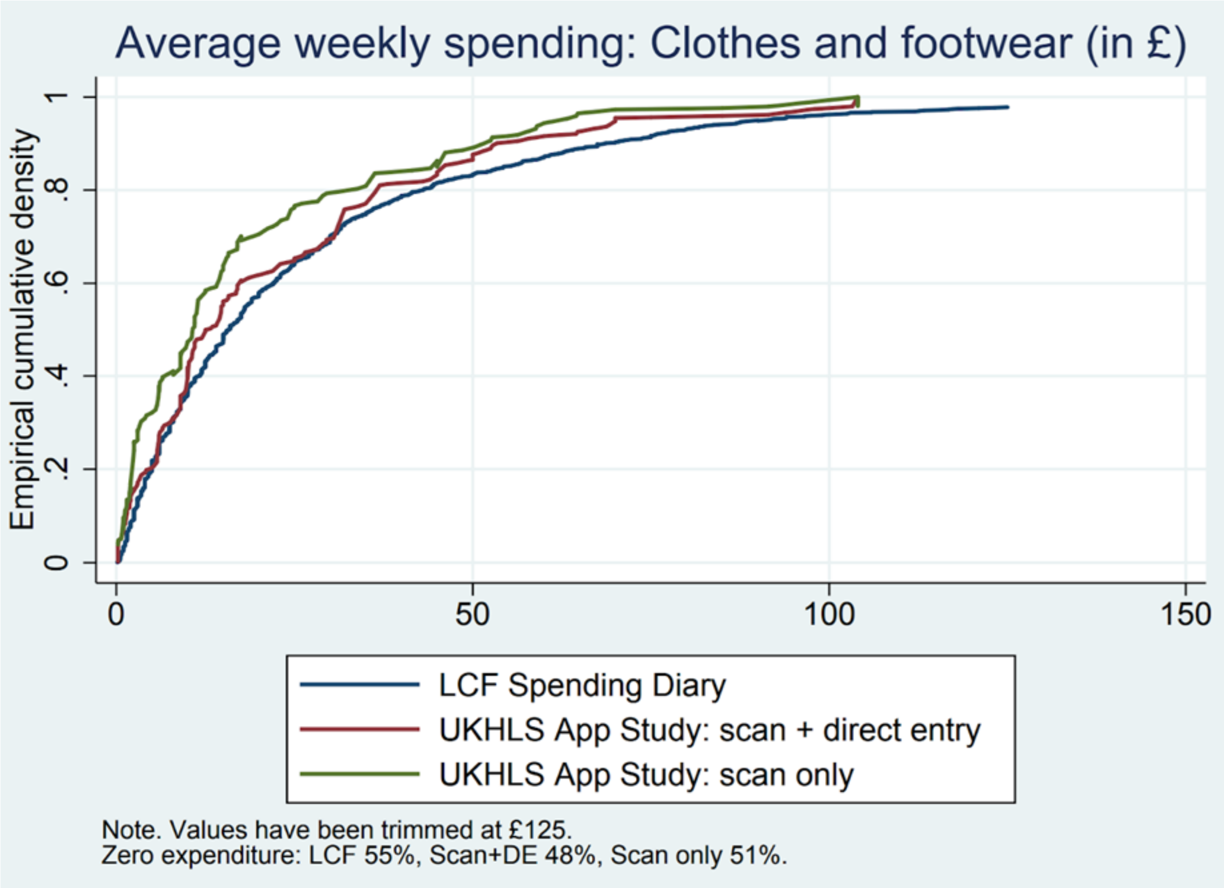
- *Daily Experiences and Well-being Study* (Fingerman et al. 2020)
- Interviewers used phone app when returning to pick device up (day 5)
- After completing all other activities, asked participant for consent to take picture of room they spend most time in
  - Up to 3 photos
  - Careful selection of motive to avoid recording any PII
- Environmental conditions of room hand-coded
  - Lighting, conditions, etc.

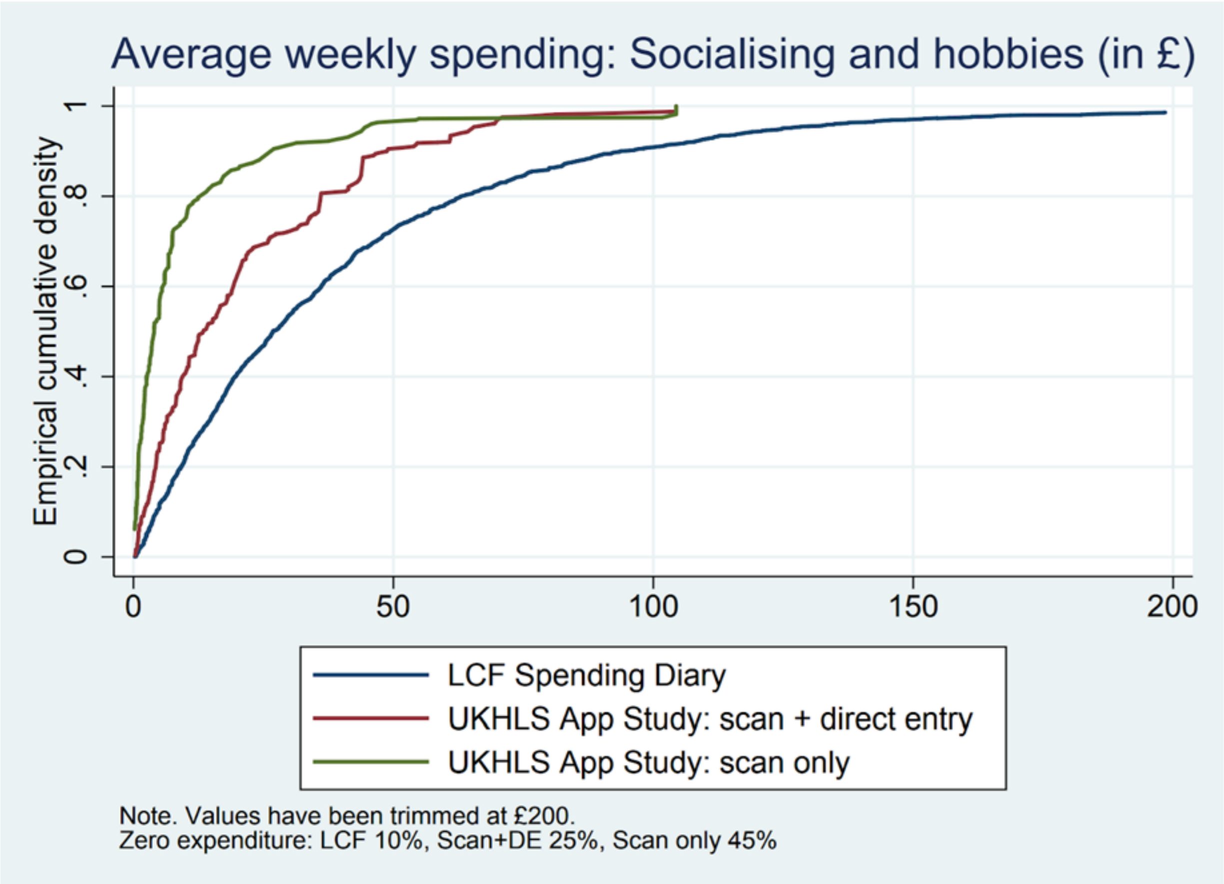


Example: How much do households spend on goods and services? (Jäckle et al. 2019; Wenz et al. 2018)

- UK Innovation Panel Wave 9 participants invited to download app (iOS & Android) to smartphone or tablet and use it to report purchases of goods and services for 1 month
  - 270 participants (13%) used app at least once
- Participant could scan and upload receipts, record purchase without receipt, report day without purchases
  - App sent push notifications once a day
- Scanned receipts hand coded
- Total expenditure (scan + direct entry) comparable to benchmark (LCF)
  - Expenditure more comparable for some categories than for others

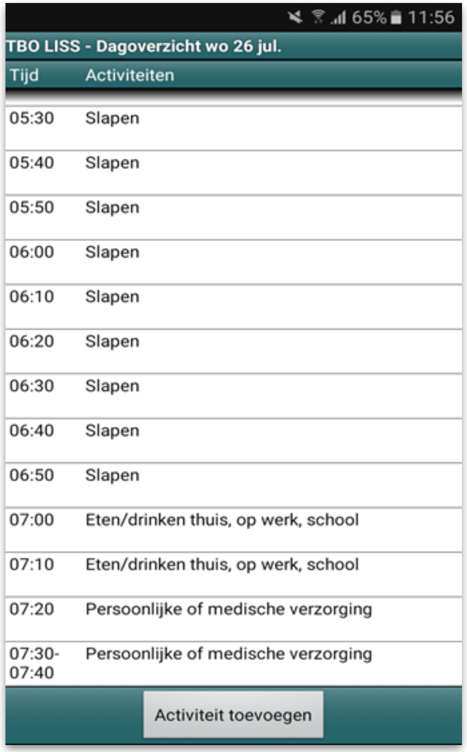
Example: How much do households spend on goods and services? (Jäckle et al. 2019; Wenz et al. 2018)





Self-reports on smartphones

- Diary studies
  - e.g., time use, expenditure, food consumption via app or web browser



Daily overview

Adding activities

Adding activity information

Elevelt et al. (2019)

Self-reports on smartphones

- Diary studies (e.g., time use, food consumption) via app or web browser
- Ecological Momentary Assessment (EMA)/Experience Sampling Method (ESM) via app
    - Collecting data several times a day on several days per week allows tracking of change within individuals in much detail
    - Immediate reporting increases ecological validity
    - Participants “pinged” to report about current circumstances
        - Objective situation: e.g., “What are you doing?”
        - Subjective state: e.g., “How anxious are you right now?”
    - Time-based vs. geolocation-based vs. event-based

Example time-based EMA: How does physical activity affect happiness? (Lathia et al. 2017)

- *Mood-Tracking Application* on smartphones of 12,000 volunteer Android users for up to 17 months
- EMA questions: affect two or more times during the day
- Physical activity for immediately preceding fifteen minute period measured both by self-report (EMA) and passively by accelerometer

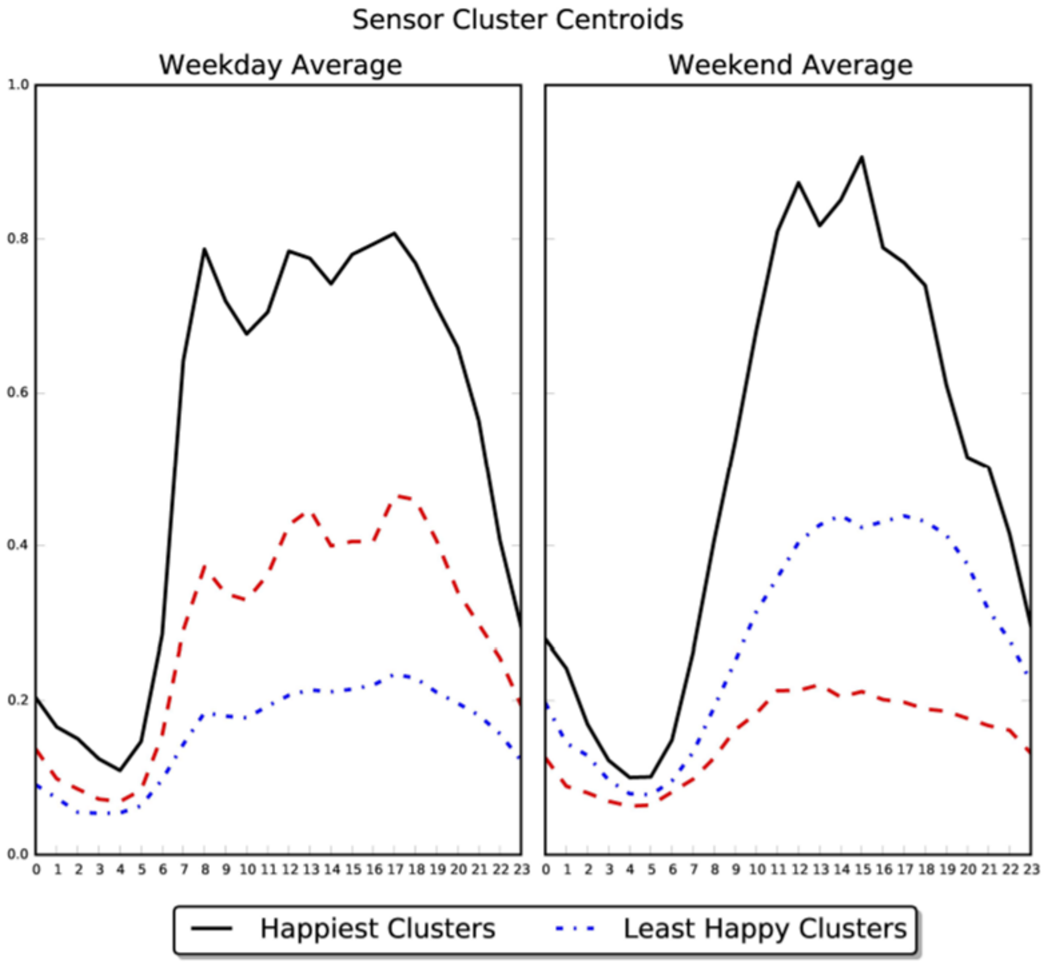
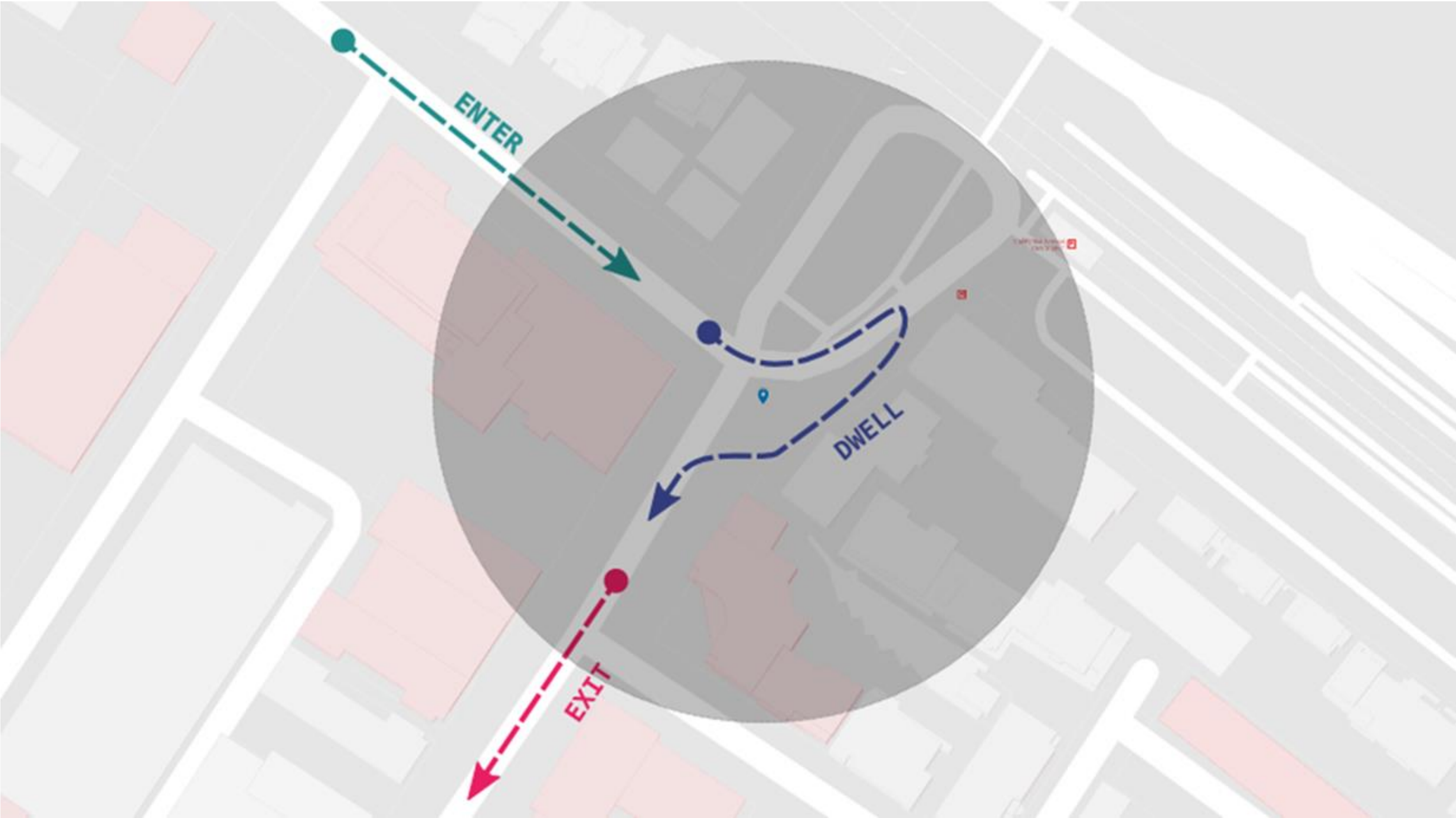


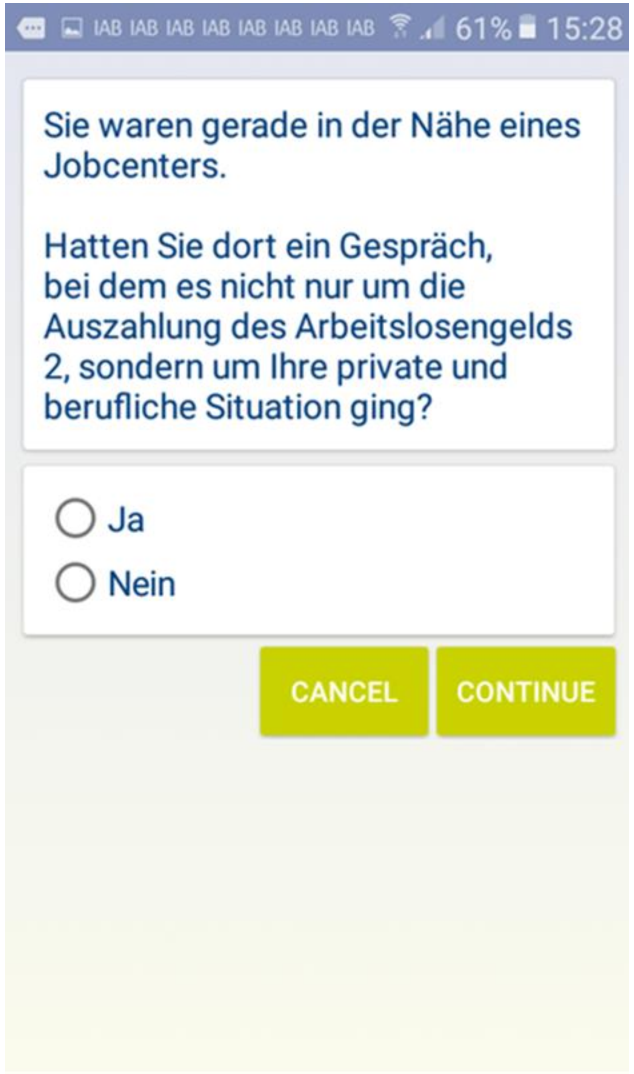
Fig 3. Centroids for the clusters generated from (left) weekday and (right) weekend activity profiles.

doi:10.1371/journal.pone.0160589.g003

Example geolocation-based EMA ("Geofencing"): Visits to job centers (Haas et al. 2020)

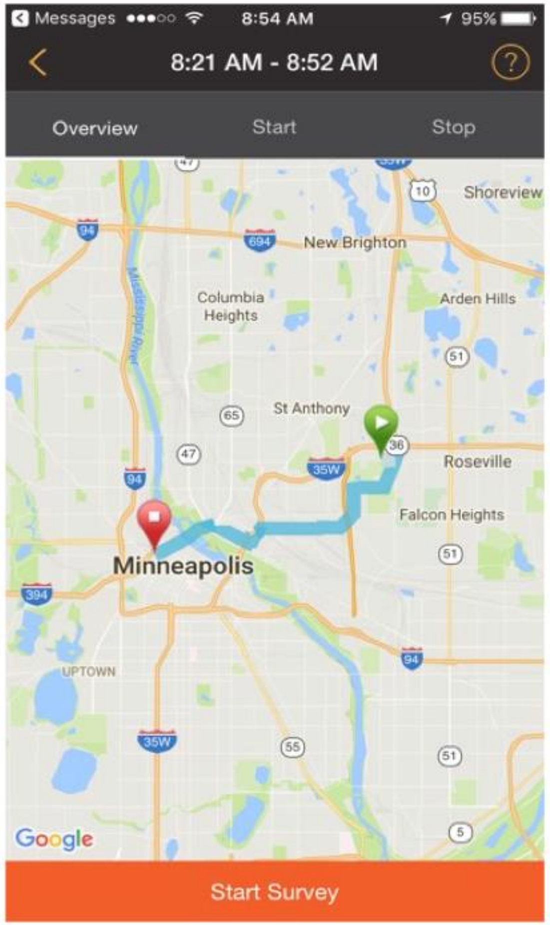


Source: https://developers.google.com/location-context/geofencing/

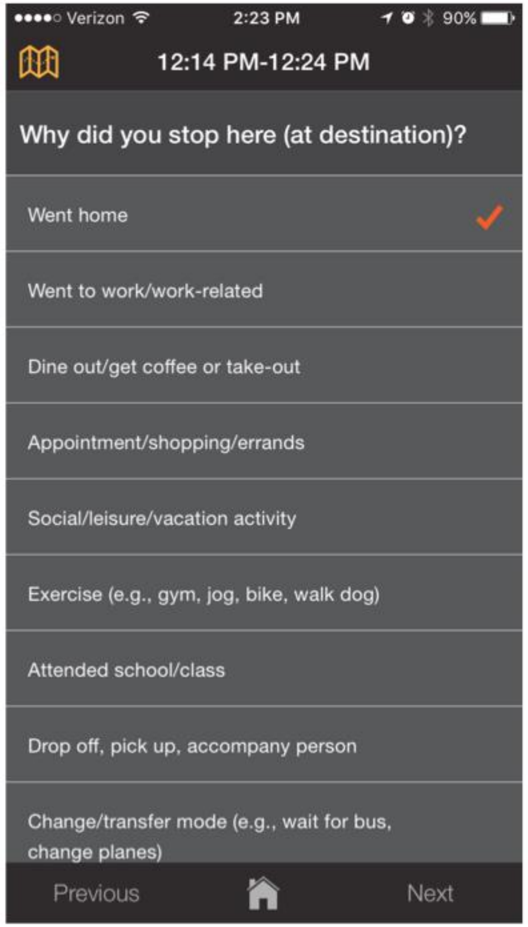


Example behavior-triggered: Trip information

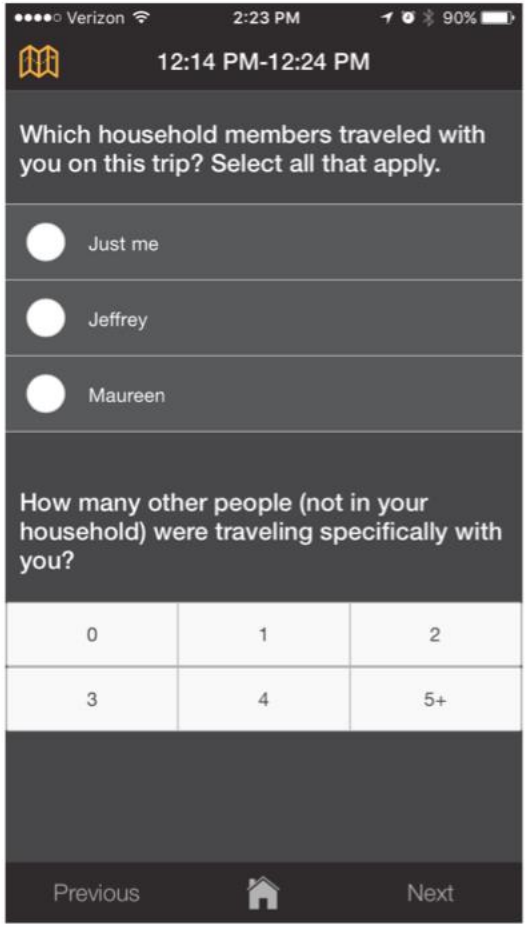
(Schmidt et al. 2021)



Trip detection



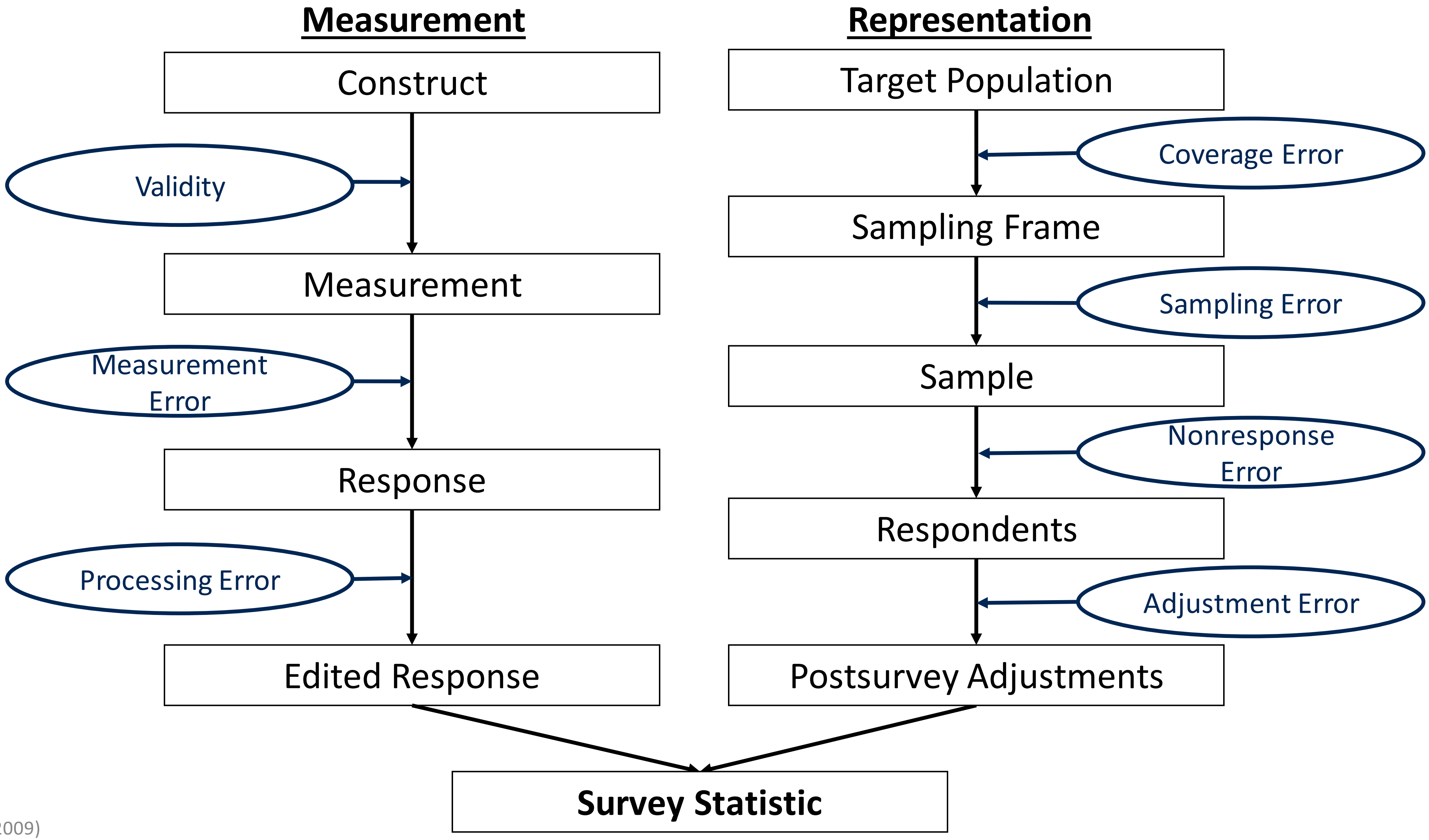
Stop purpose



Travel companions

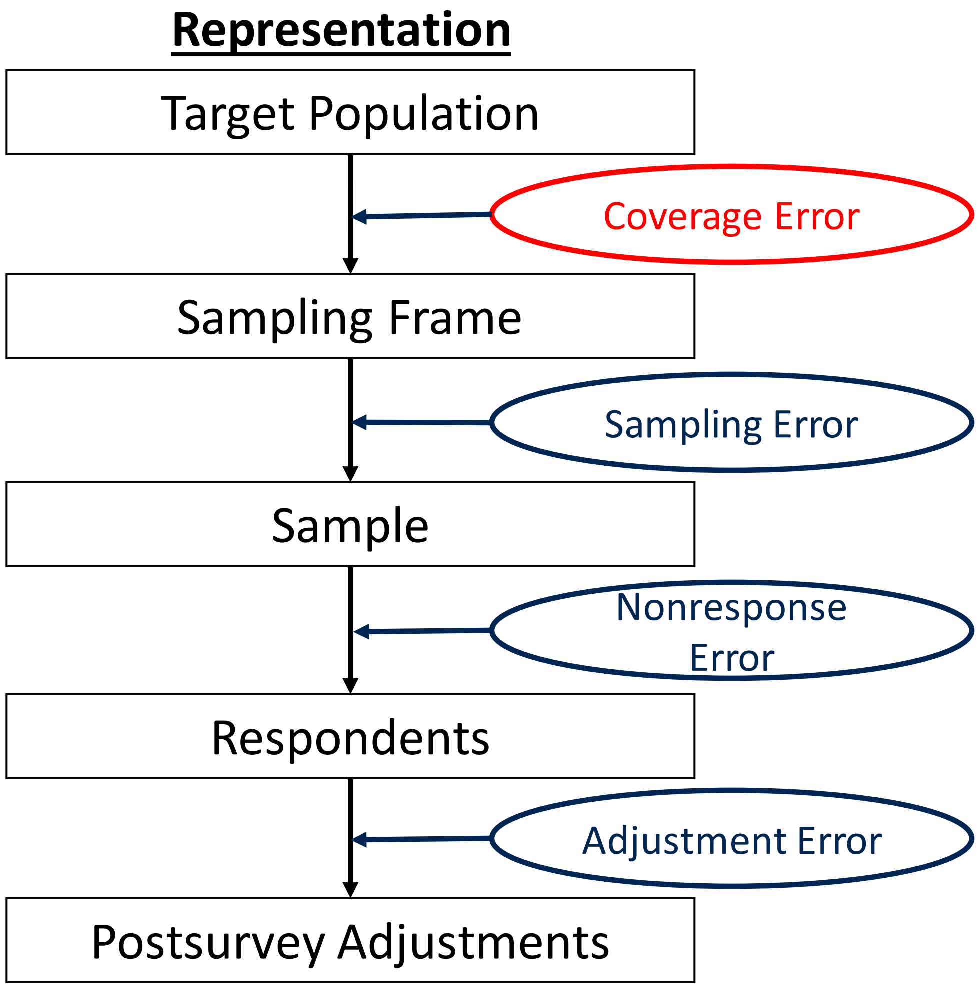
Research & study design

Total survey error (TSE) framework



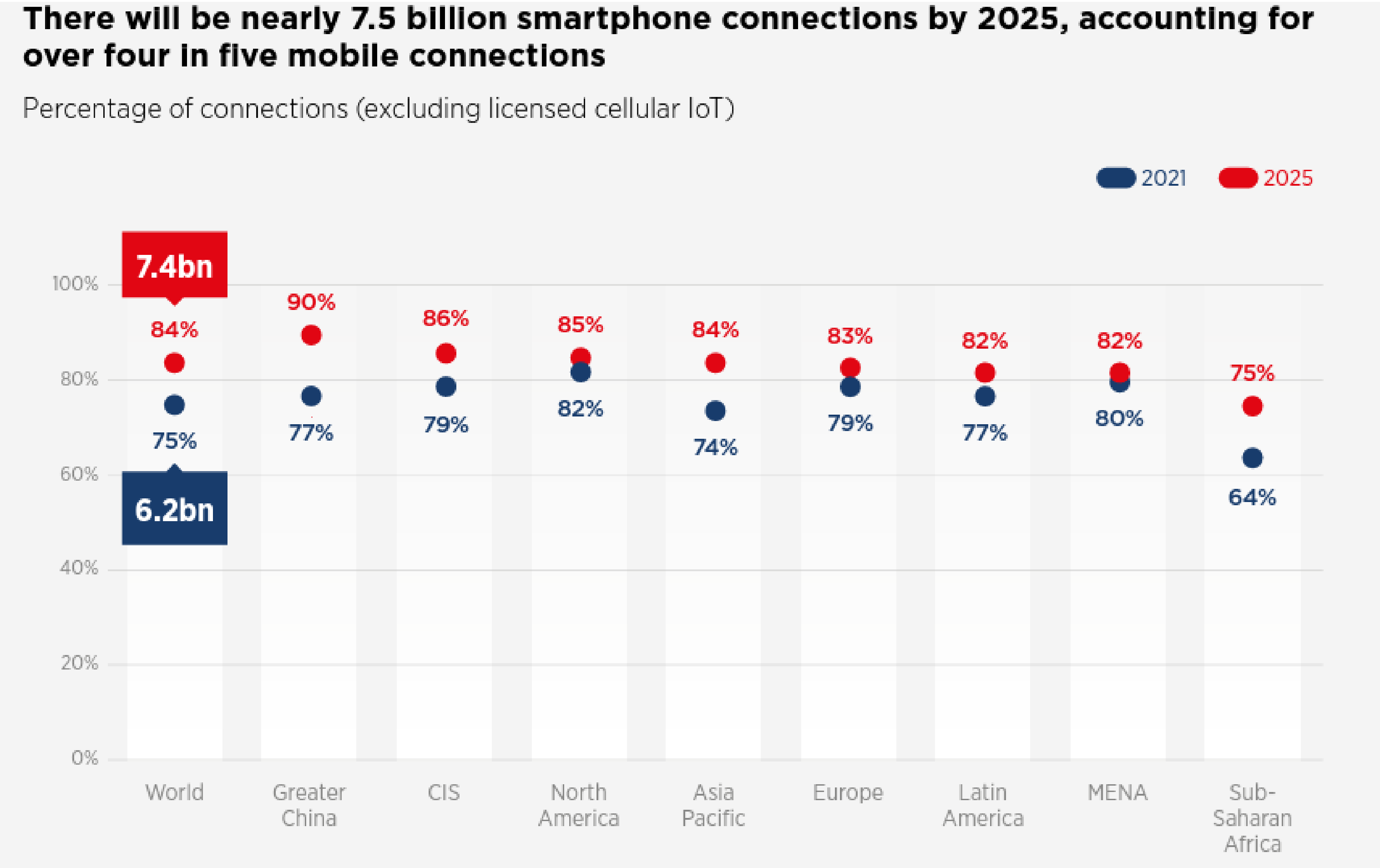
Groves et al. (2009)

Representation error in app, sensor & wearables data collection



- **Coverage error:** A study relies on participants to share data from their fitness wristbands to analyze weekend vs. weekday activity by race & ethnicity. The rate of ownership of these devices is lower in the study population than in the general population.

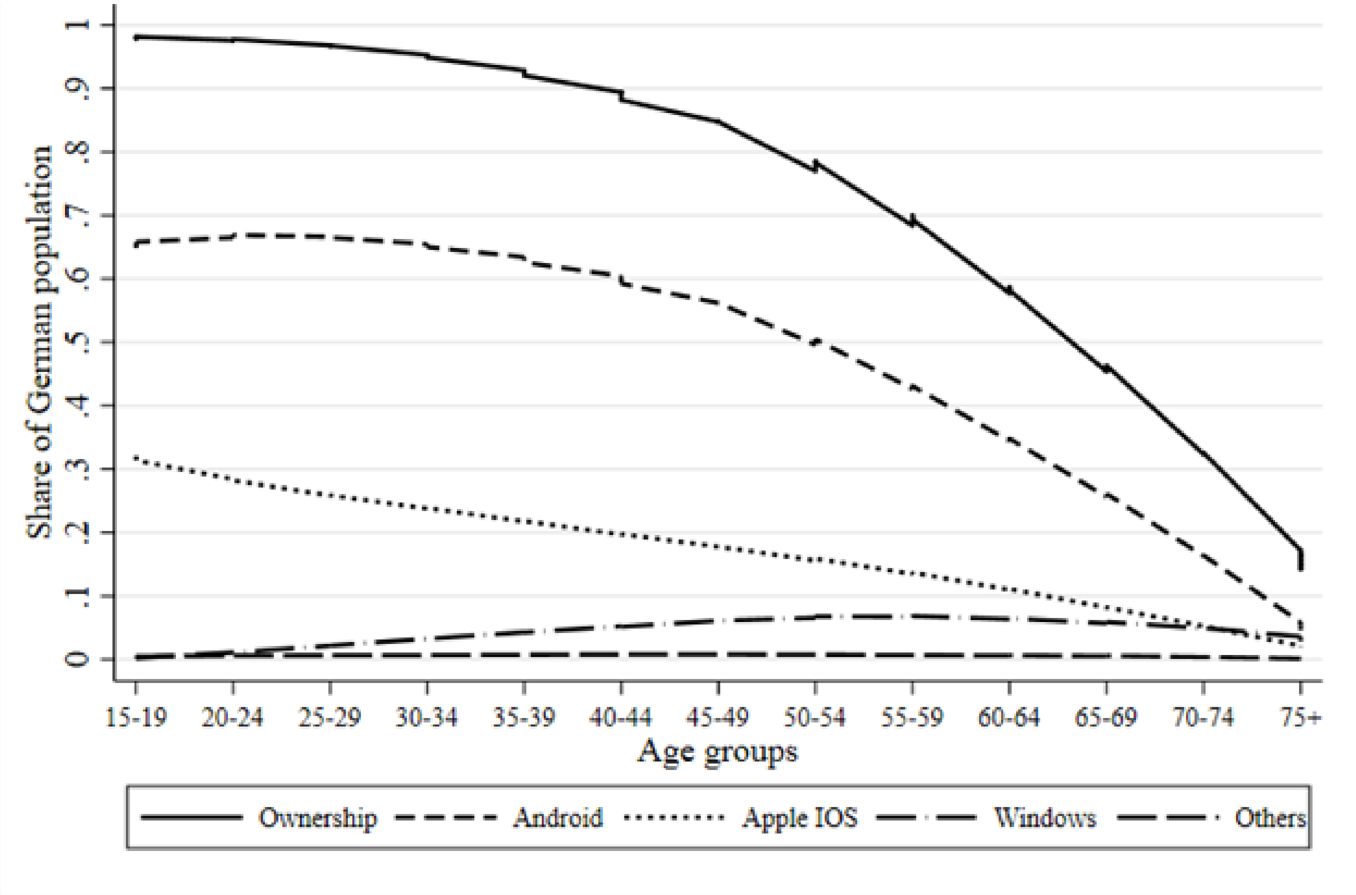
BYOD: Coverage smartphones



Source: https://www.gsma.com/mobileeconomy/wp-content/uploads/2022/02/280222-The-Mobile-Economy-2022.pdf

BYOD: Smartphone coverage bias in Germany

(Keusch et al. 2020)



- Smartphone ownership higher among…
  - …younger
  - …male
  - …higher educated
  - …people in New States
  - …people living in larger communities

➢ Digital Divide

Source: PASS Wave 11; n = 13,703; Locally weighted scatter-plot smoother (LOWESS) regression

BYOD: Smartphone coverage bias in Germany

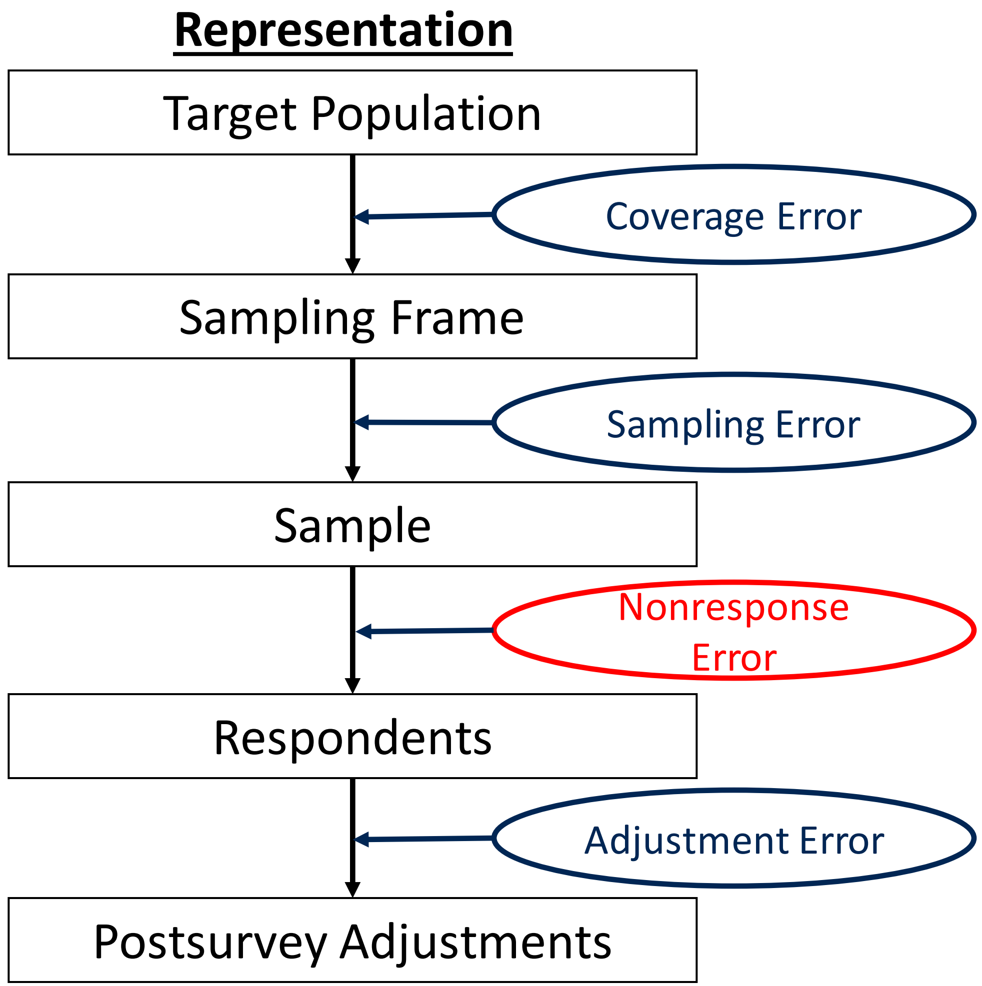
(Keusch et al. 2020)

- *Overall smartphone coverage bias* in many substantive estimates relatively small; especially once adjusting for sociodemographic differences between general population and smartphone owners
    - High social inclusion: +2.8 p.p.
    - Size of personal network: n.s.
- Comparable *Android smartphone coverage bias* after sociodemographic adjustment
    - High social inclusion: +1.6 p.p.
    - Size of personal network: n.s.
- Much larger *iPhone coverage bias*, even after adjusting for sociodemographics (up to 11 p.p.)

Solution to coverage problem: Provide (loaner) devices

- Provide participants with device and ask to share data for long durations of time
- Examples:
    - General population (e.g., Scherpenzeel 2017)
    - Older adults (e.g., Compernolle et al. 2022; English et al. 2022; Fingerman et al. 2020, 2022; Fritz et al. 2017; Huo et al. 2020; Maher et al. 2018; York Cornwell & Cagney 2017, 2020)
    - Chronically ill (e.g., Goodspeed et al. 2018)
    - Recent parolees (e.g., Sugie 2018)
    - Students (e.g., Wang et al. 2014)

Representation error in app, sensor & wearables data collection



- **Nonparticipation error:** Participants are provided with actigraphs to measure sleep patterns for a week. Those who do not sleep well remove the device at night because it disturbs their sleep.

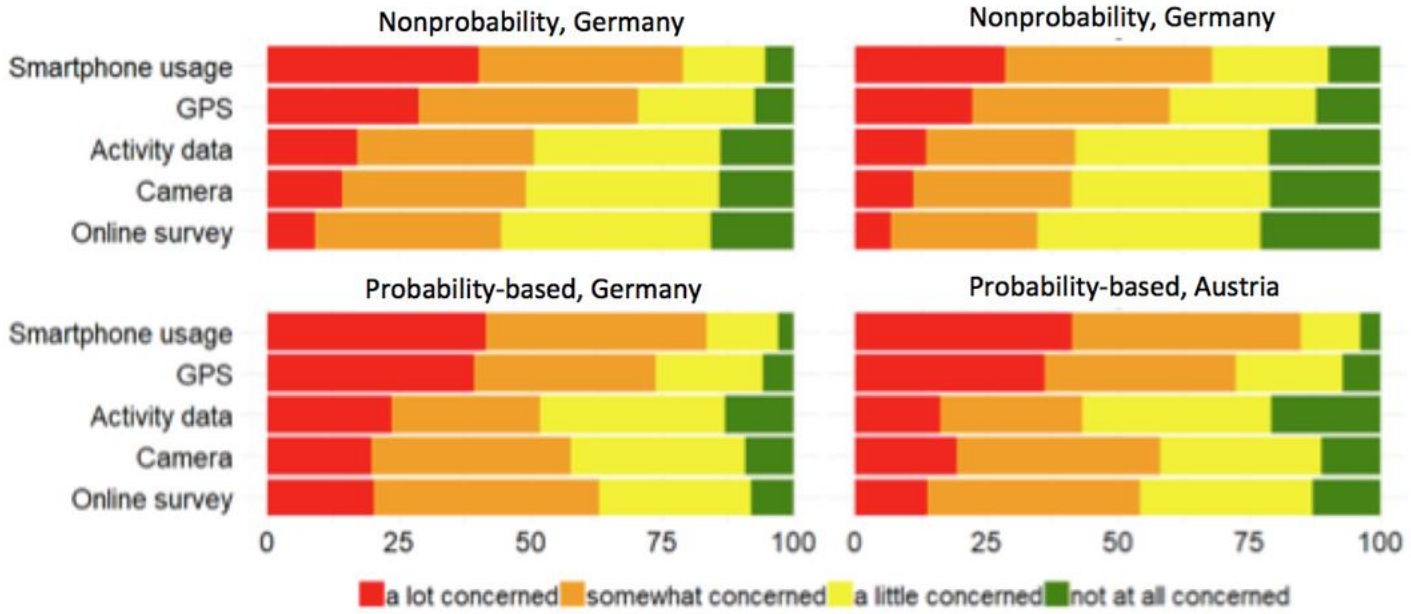
Nonparticipation: Willingness to participate (WTP) & actual participation

- Many studies work with self-selected volunteer samples
- General willingness varies by sensor and task
  - 29% (Spain) - 52% (Mexico) for taking pictures, 19 (Portugal) - 37% (Chile) for sharing GPS location (Revilla et al. 2016)
  - Mobility (GPS, accelerometer): 37% willing (81% participate) (Scherpenzeel 2017)
  - Physical activity (wearables): 57% willing (90% participate) (Scherpenzeel 2017)
- Willingness/Downloading research app in general population usually rel. low
  - 35% downloaded & registraterd CBS Travel app (McCool et al. 2021)
  - 24% registered in the Household Budget app (Rodenburg et al. 2022)
  - 18% would install app to track URLs of visited websites (Revilla et al. 2019)
  - 17% downloaded UK Understanding Society IP budget app (Jäckle et al. 2019)
  - 15% downloaded IAB-SMART app (Keusch et al. 2022a)

Mechanisms of (non-)participation: Privacy concern

- Participants might have concerns about potential risks related to sensor data
  - Data streams could be intercepted by unauthorized party
  - Connecting multiple streams of data could re-identify previously anonymous users
  - Information could be used to impact credit, employment, or insurability
- Higher privacy & security concerns correlate with lower WTP

(Keusch, et al. 2019; Revilla et al. 2019; Struminskaya et al. 2020; 2021; Wenz et al. 2019; Wenz & Keusch in press)



Keusch et al. (2021)

Mechanisms of (non-)participation: Incentives

- Inconsistent findings of effect of incentives on participation
- Hypothetical WTP increases with incentives for downloading app and staying until end of study (Keusch et al. 2019; Wenz & Keusch in press)
- IAB-SMART, Germany (Haas et al. 2021)
  - 20€ for installation increase installation rate over 10€ (16% vs. 13%)
  - Bonus incentive for consenting to all 5 data collection functions no effect
- Statistics Netherlands Travel App (McCool et al. 2021)
  - 5€ unconditional + 5€ Registration + 5€ after 7 days: 30%
  - 5€ unconditional + 10€ after 7 days: 36%
  - 5€ unconditional + 20€ after 7 days: 40%
- UK IP Spending Study (Jäckle et al. 2019)
  - £6 incentive for installation does not increase installation rate over £2

Other Mechanisms on (non-)participation

- Agency: WTP higher for tasks where participants have agency over data collection (Revilla et al. 2019; Keusch et al. 2019; Struminskaya et al. 2020; 2021; Wenz & Keusch in press)
- Sponsor: WTP higher for university sponsor vs. market research and statistical office (Keusch et al. 2019; Struminskaya et al. 2020)
- Framing: emphasizing benefits does not influence WTP (Struminskaya et al. 2020; 2021)
- Smartphone skills: more activities on smartphone (e.g., using GPS, taking pictures, online banking, etc.) correlates with higher WTP (Keusch et al. 2019; Struminskaya et al. 2020; 2021; Wenz et al. 2019; Wenz & Keusch in press)
- Experience: prior research app download increases WTP (Keusch et al. 2019; Struminskaya et al. 2020; 2021)
- Sociodemographics: educational attainment (Jäckle et al. 2019; Keusch et al. 2021, 2022; McCool et al. 2021; Wenz & Keusch in press) and age (Jäckle et al. 2019; McCool et al. 2021; Keusch et al. 2022; Wenz & Keusch in press) correlated with WTP

Non-participation for (loaner) wearables (Actigraphy, Fitbit, etc.)

- Must consider both *Consent Rate* and *Compliance Rate*
- Compliance can be for full study duration (all days, all hours) or partial (some days, some hours)
- Reasons for non-participation include:
  - Device not visually appealing
  - Device uncomfortable
  - Device removed (at night or to shower) and not put back on
  - Battery runs out
  - Data does not sync (calibration error, syncing error)
  - Device lost
  - Device not returned
  - ...

Toepoel (2021)

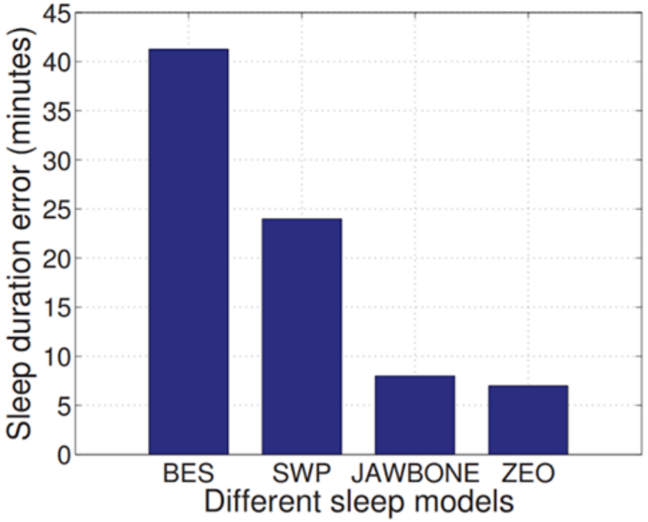
Can you really measure sleep using sensors?

- Does physical inactivity and low heart-rate equal sleep?

Percent of time that self report differs from accelerometer	
Self_report: asleep; Accelerometer: not in bed	6.8%
Self_report: not asleep; Accelerometer: in bed	14.9%
Self_report: asleep; Accelerometer: not asleep	14.0%
Self_report: not asleep; Accelerometer: asleep	10.5%

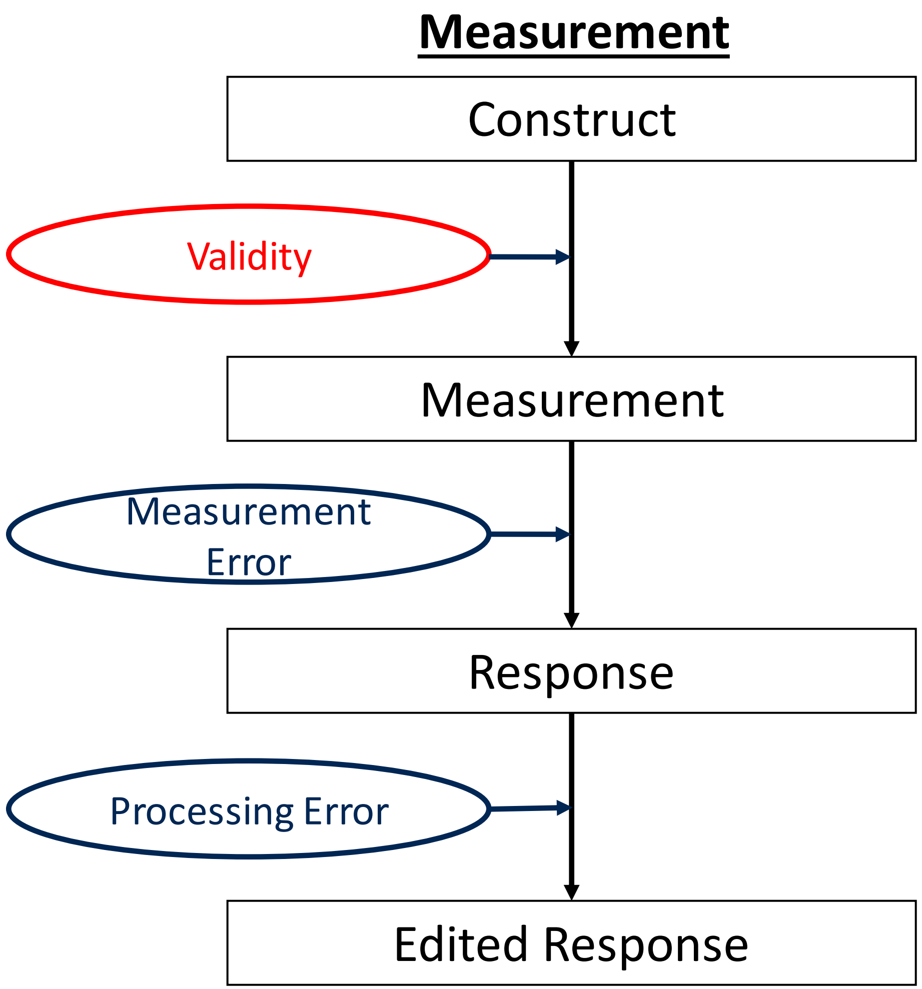
Kapteyn et al. (2019)

- Does absence of light, sound, and activity measured by a smartphone equal sleep?
- But for <u>some phenomena</u>, sensors seem to be provide highly valid data



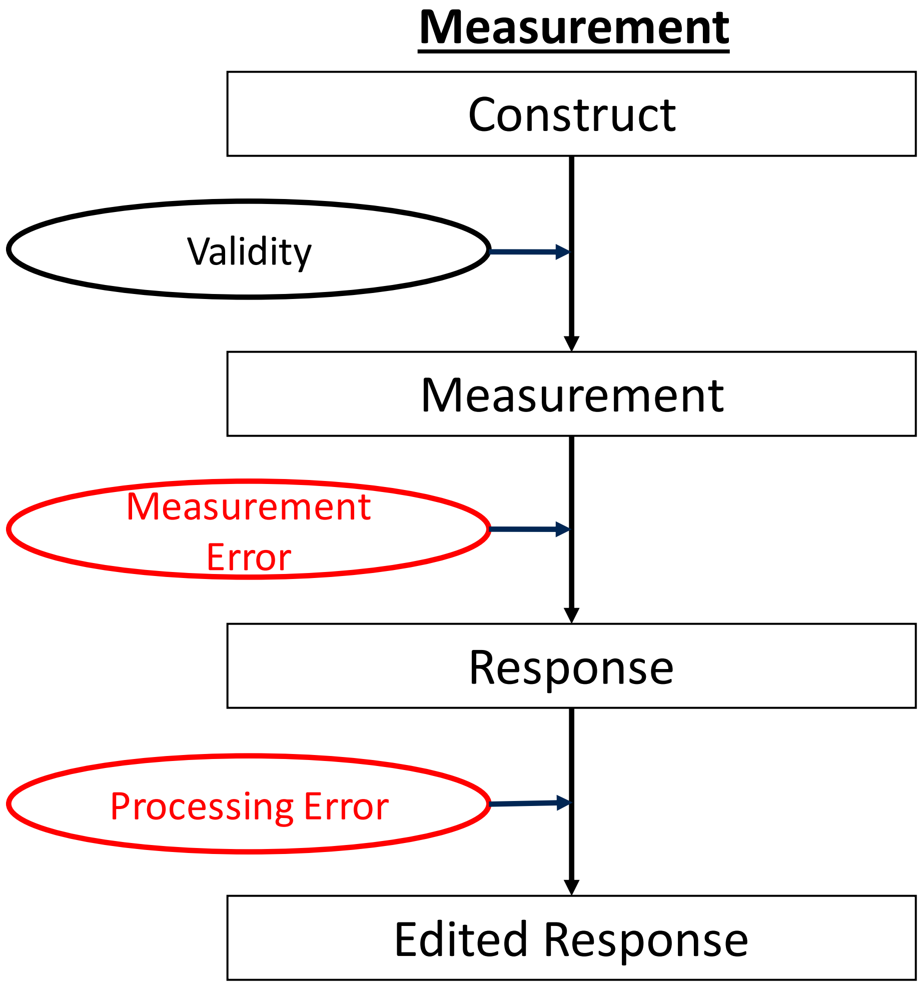
Chen et al. (2013)

Measurement error in app, sensor & wearables data collection



- **Validity:** You are using actigraphy to detect intensity of physical activity in a sample of older adults. Your study population is very sedentary and it is difficult to identify physical activity versus usual activity.

Measurement error in app, sensor & wearables data collection



- **Measurement error:** GPS is less precise in urban areas where there are many large buildings.
- **Processing error:** Raw accelerometer data is classified as different types of activity based on where sensor/phone is located (e.g., pocket vs. purse).

Practical implementations and operational considerations

Practical implementations and operational considerations

1. Participant recruitment
2. Consent
3. Frequency of sensor measurement
4. Battery life
5. Storage & data transfer
6. Costs

Participant recruitment

- Volunteers/enthusiasts recruited through social media, online ads, etc.
  - e.g., *Mappiness* app downloaded by ~22,000 users (MacKerron & Mourato 2013)
- Members of non-probability online access panels
  - e.g., as part of a survey (Revilla et al. 2016)
- Special-interest groups recruited through flyers, face-to-face, telephone, etc.
  - e.g., students (Wang et al. 2014; Stopczynski et al. 2017), older adults (Fingerman et al. 2019, 2022; Fritz et al. 2017; Hou et al. 2020; Maher et al. 2018; York Cornwell & Cagney 2017, 2020), parolees (Sugie 2018)
- General population
  - Participants of existing probability-based panels: e.g., LISS Panel (Scherpenzeel 2017), UK Understanding Society Innovation Panel (Jäckle et al. 2019), German PASS Panel (Kreuter et al. 2020)
  - From population registry (McCool et al. 2021)

Example recruitment: IAB-SMART
(Kreuter et al. 2020)

- Sampling from PASS panel participants aged 18-64
- Wave 11 in 2017:
  - Do you own a smartphone?: 84% YES
  - Which operating system do you use?: 70% Android
- Smartphone owners with Android operating system
  - Access to required sensor data only possible through Android OS
- Recruitment in Jan/Feb 2018 via mail with one reminder

Example recruitment: IAB-SMART

(Kreuter et al. 2020)

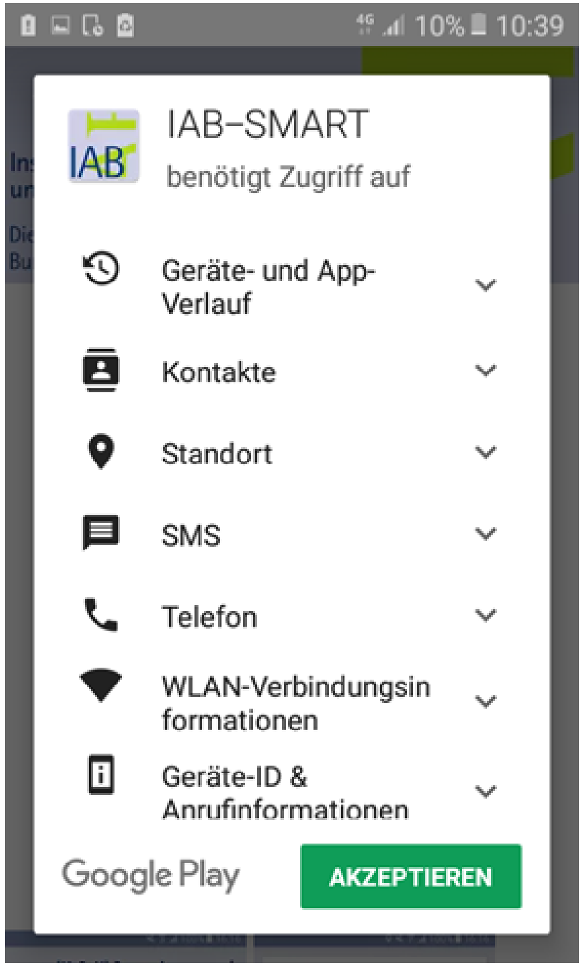
- Invitation package
  - Cover letter
  - Data protection information
  - Voucher flyer
  - Installation booklet
- www.iab.de/smart
  - Frequently asked questions
- E-Mail address & telephone hotline

Consent

- Ethics
  - “Informed consent”
  - IRB approval
- Legal
  - Depends on type of data collected (e.g., PII, GDPR, DPIA)
  - Talk to legal department early!
- Technical implementation
  - Depends on device, OS, and researcher choices

Example consent: IAB-SMART (Kreuter et al. 2020)

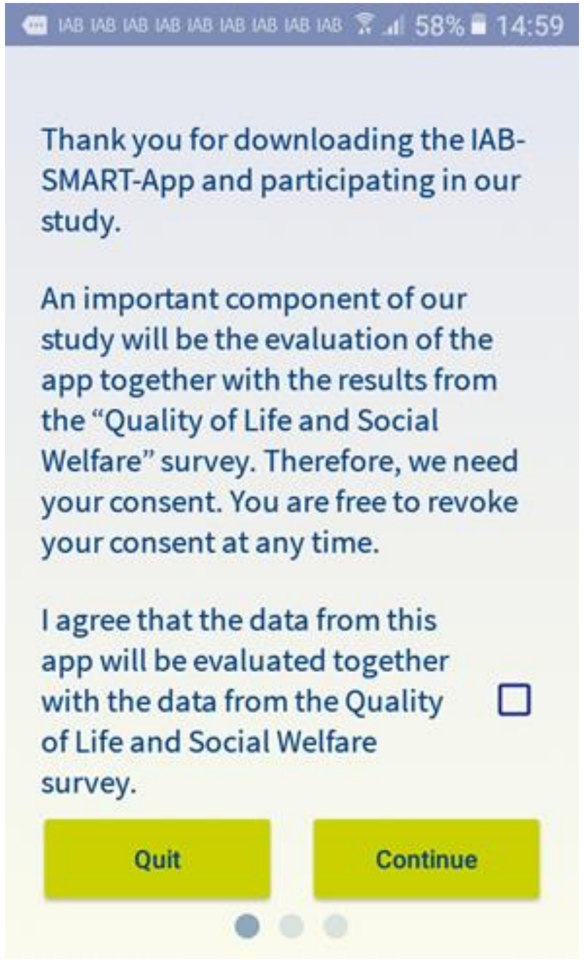
Google Play Store



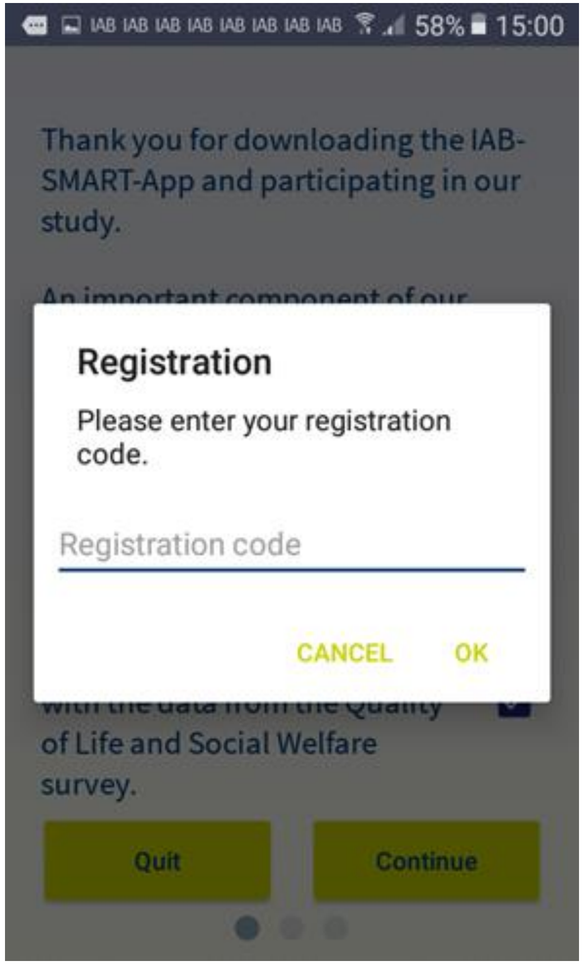
Google Permissions

App installed

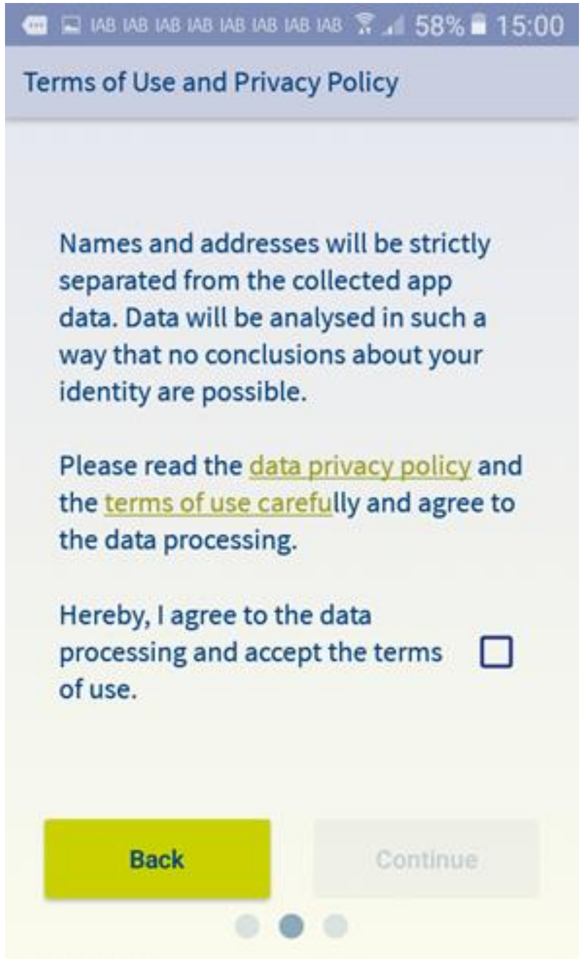
Example consent: IAB-SMART (Kreuter et al. 2020)



Linkage to PASS



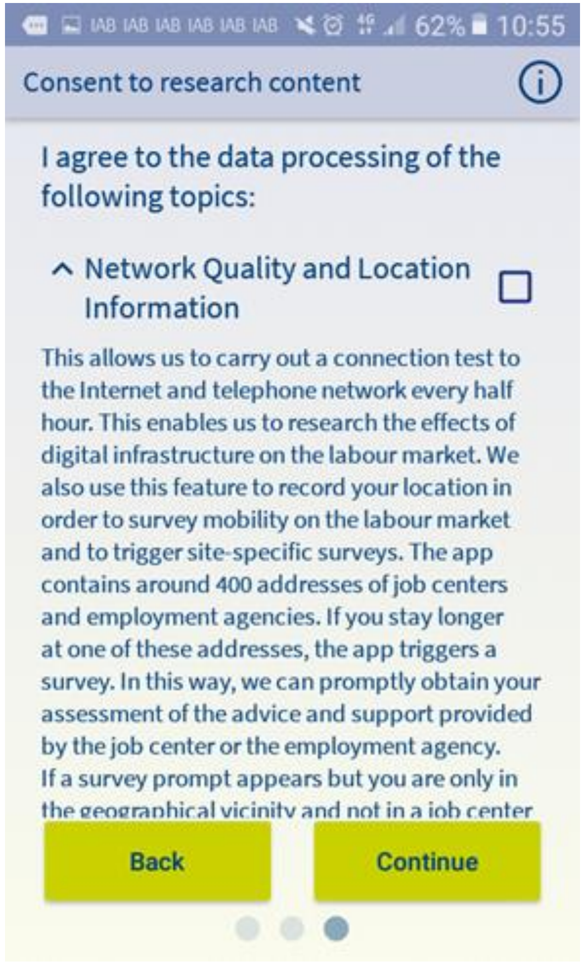
Registration



Data processing

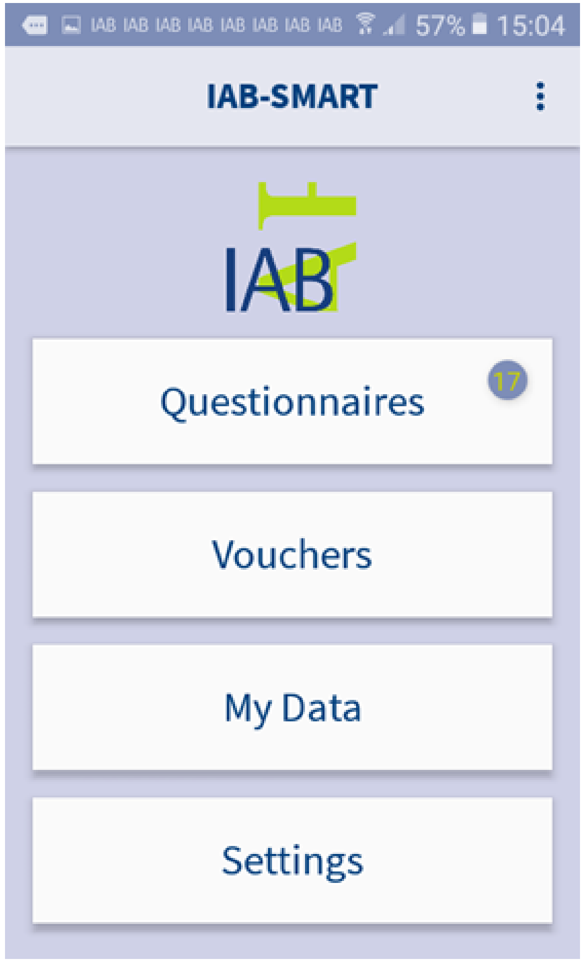
Example consent: IAB-SMART (Kreuter et al. 2020)

Individual consent screen



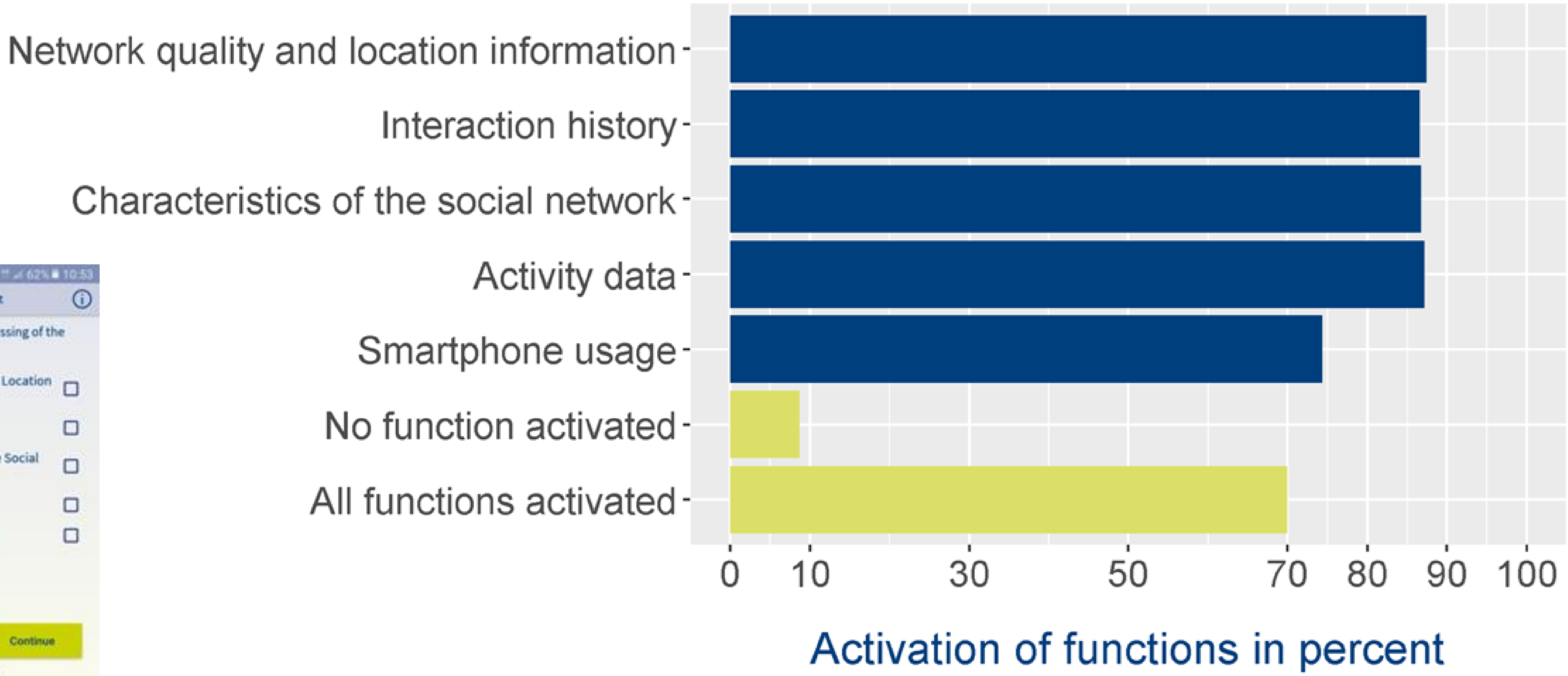
Function explanation

Full consent

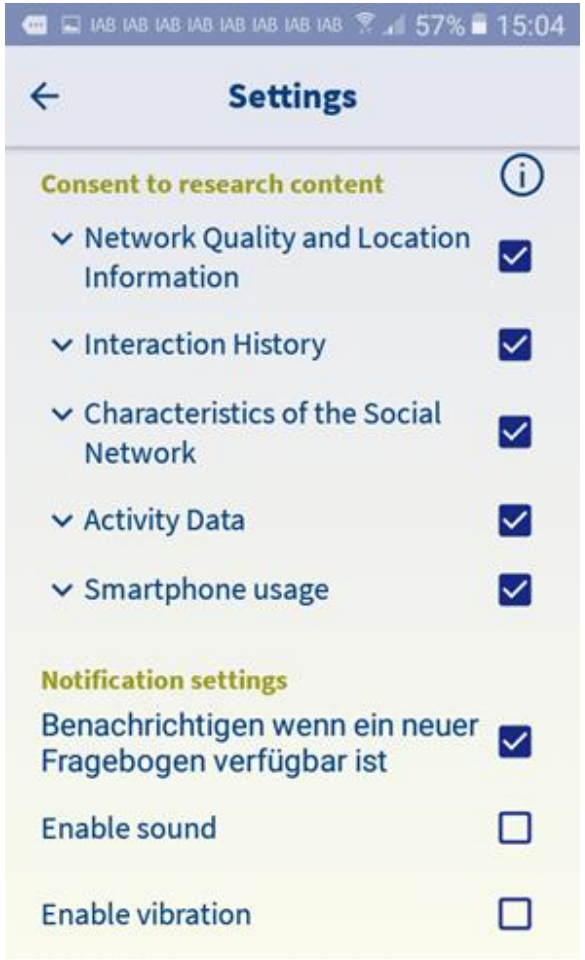


App home screen

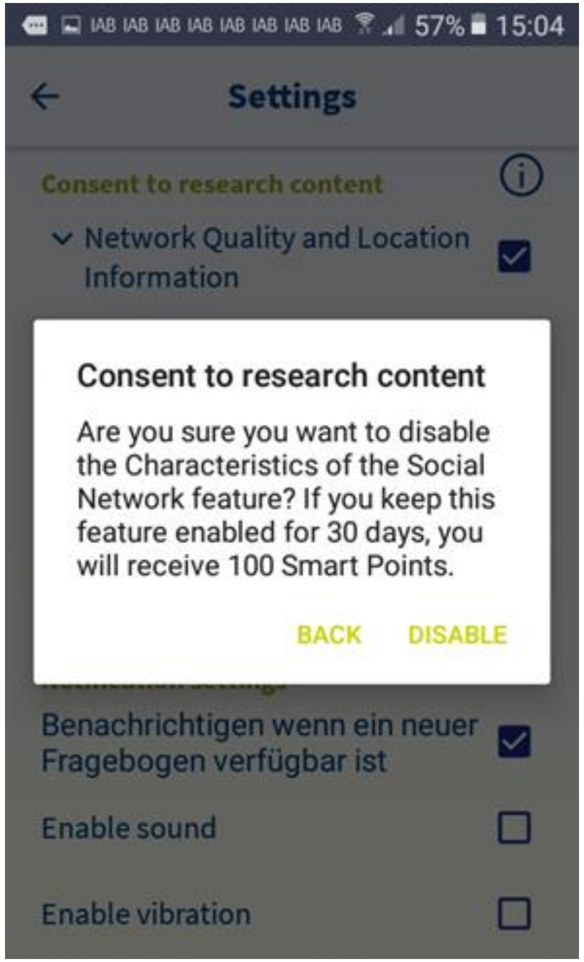
Example consent: IAB-SMART (Kreuter et al. 2020)



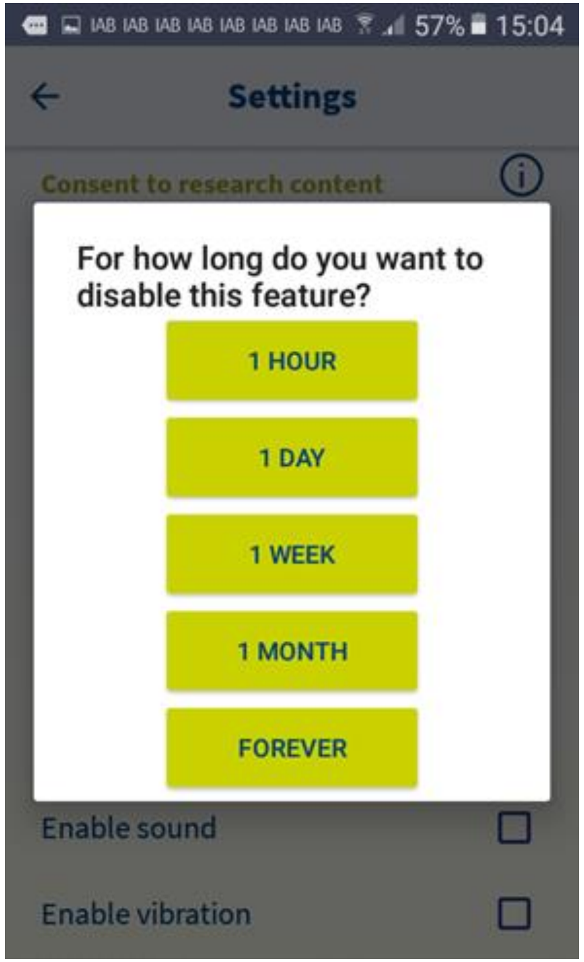
Example consent: IAB-SMART (Kreuter et al. 2020)



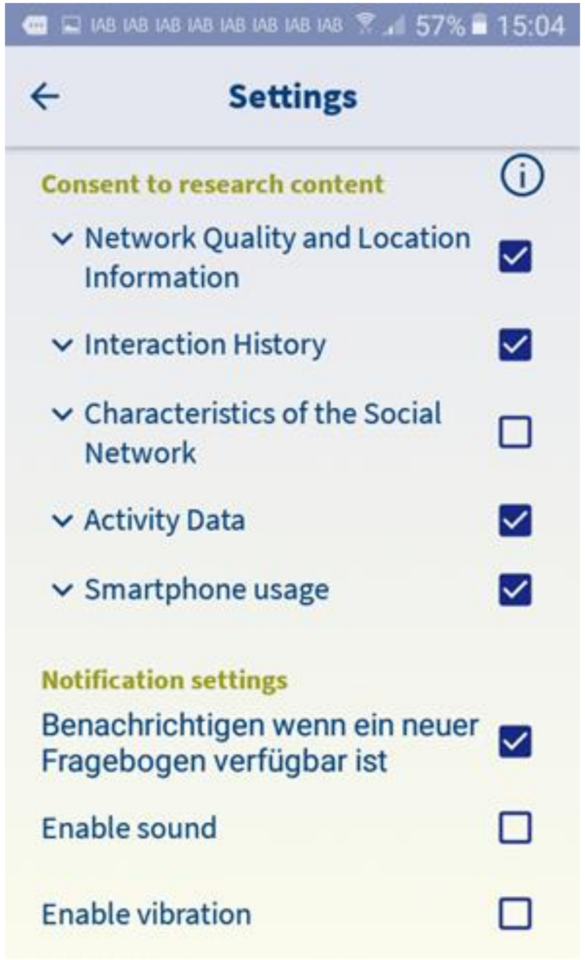
App settings



Withdrawing consent



Withdrawing consent



App settings

Example consent: One-time GPS (Struminskaya et al. 2021)

General consent

Framing, agency, privacy explanation

GPS measurement

Frequency of sensor measurement

- *Sampling rate* defines frequency of measurement (usually in Hz)
- Realized frequency depends on various factors
  - Sensor's technical capabilities (max. sampling rate)
  - Outside factors (e.g., sleep/battery saving mode, technical failure)
  - Design decisions by researcher
- In practice, there seem to be four groups of measurement frequencies
  - Contingent
  - Discrete
  - Continuous
  - Combination

Frequency of sensor measurement

- Contingent
    - Measurement only at specific times, adding individual data points to survey, as if additional question was asked
    - e.g., GPS location whenever EMA is answered (MacKerron & Mourato 2013)

Frequency of sensor measurement

- Contingent
- Discrete
  - Usually to conserve battery and storage and/or to protect privacy
    - In case of GPS, allows to calculate activity radius but not specific traces
  - e.g., GPS every 5 min from 9 am to 9 pm (York Cornwell & Cagney 2017), every 15 minutes (Sugie 2018), every 30 minutes (Kreuter et al. 2020)
  - e.g., audio recordings for 30s every 7 min during waking hours (Fingerman et al. 2020, 2022)

Frequency of sensor measurement

- Contingent
- Discrete
- Continuous
  - Tracking of smartphone-mediated behavior usually always on (e.g., Kreuter et al. 2020; Stachl et al. 2020; Sugie 2018)
  - Accelerometer & gyroscope (and some others) usually always on to detect movement
    - Actual frequency depends on device, model, etc.
  - GPS collected at high frequency allows measurement of exact route
    - But may have negative effect on phone performance
  - Microphone always on (e.g., Wang et al. 2014)
    - But data pre-processed on device to save storage and preserve privacy

Frequency of sensor measurement

- Contingent
- Discrete
- Continuous
- Combination
  - Combining fine grained tracking with saving battery and reducing invasiveness
  - e.g., measure activity (accelerometer) only at specific times during the day - 15 min twice a day (Lathia et al. 2017)
  - e.g., reduce sampling rate of GPS if idle, based on accelerometer measures - once every second when in motion but only every minute when still (McCool et al. 2021)

Battery life

- High sampling rates might reduce battery
  - Especially for GPS tracking

- Newer models might go to sleep mode and/or turn off data collection when reaching low battery level

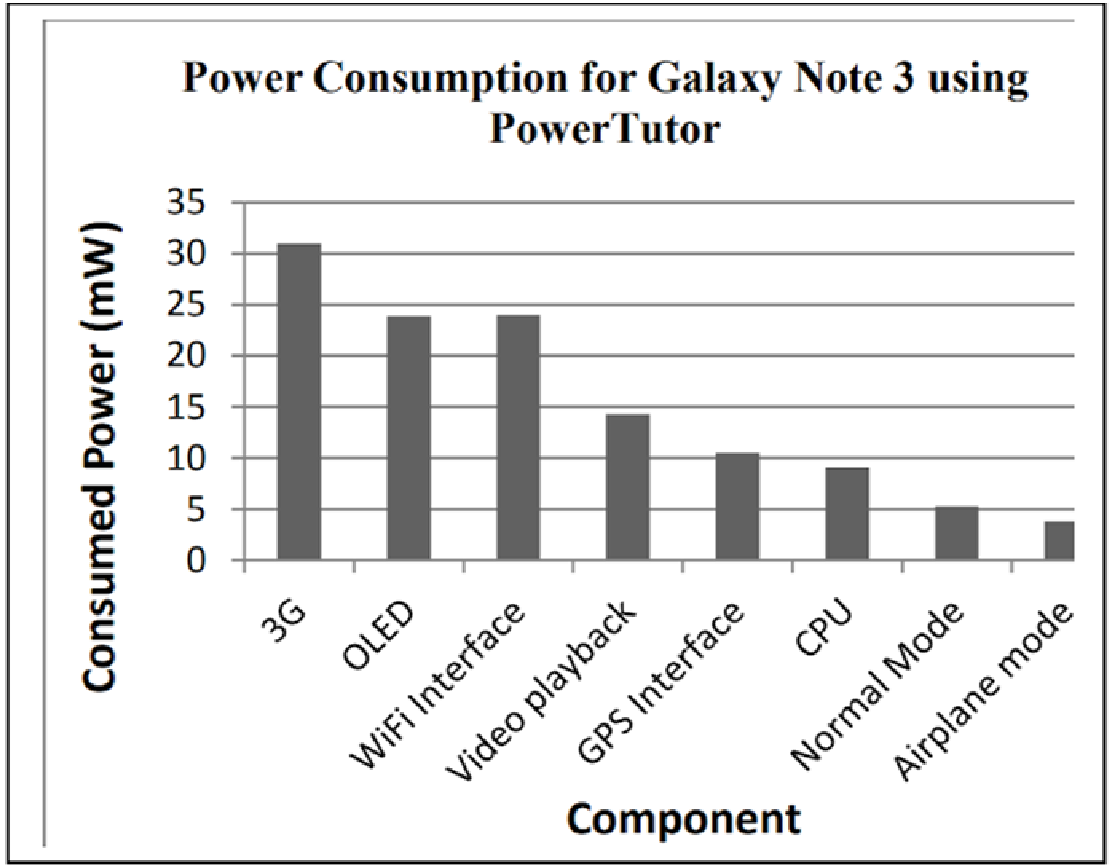
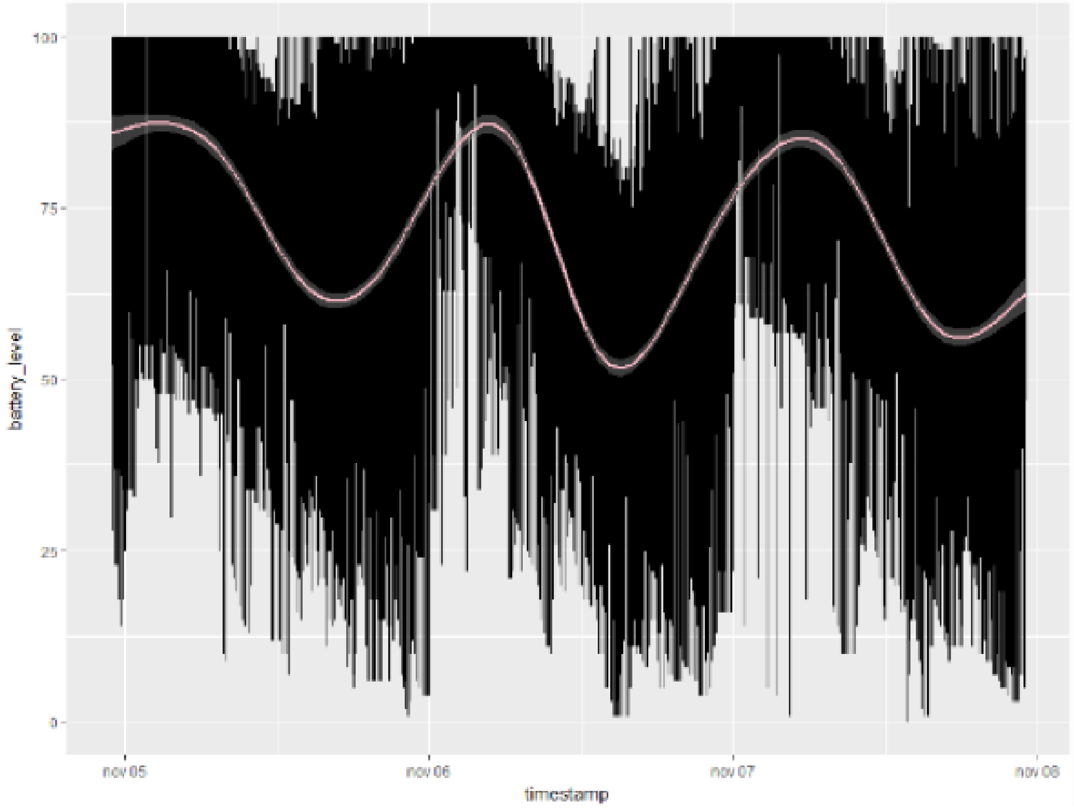


Fig. 5. Power measurements for Galaxy Note3 using PowerTutor.

Tawalbeh et al. (2016)

Battery life example: Statistics Netherlands Travel App (McCool et al. 2019)

- Battery levels for all participants Nov 5-8, 2018
  - Battery levels follow circadian pattern
  - Very few batteries run empty over course of four days



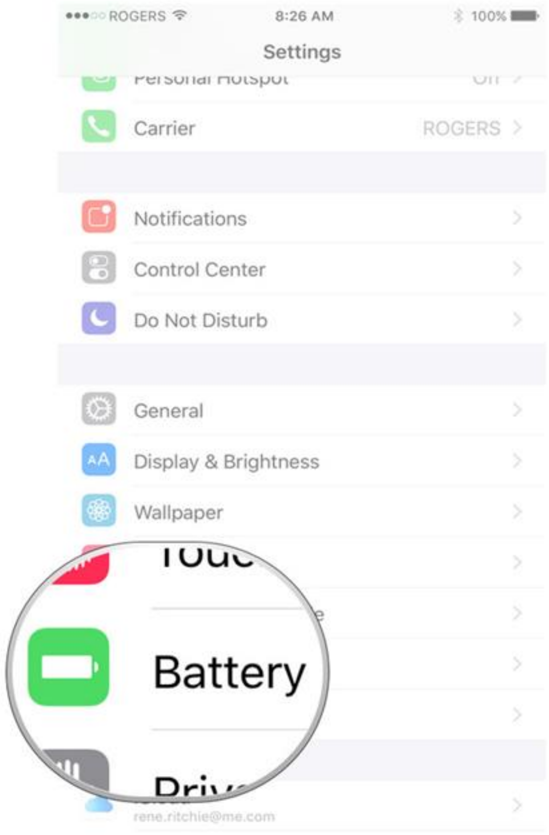
- For loander devices, participants need to be reminded to charge battery

Exercise

Find out what drains your battery!

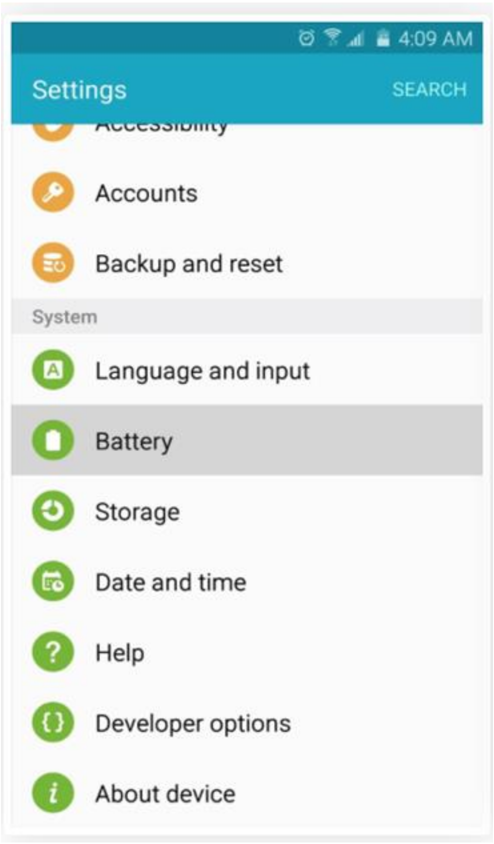
- Settings → Battery

iPhone



Source: [https://www.imore.com/how-see-whats-using- battery-life-your-iphone-or-ipad](https://www.imore.com/how-see-whats-using-battery-life-your-iphone-or-ipad)

Android



Source: [https://android.gadgethacks.com/how-to/whats- draining-your-androids-battery-find-out-fix-for-good-0162267/](https://android.gadgethacks.com/how-to/whats-draining-your-androids-battery-find-out-fix-for-good-0162267/)

Data transfer

- Some systems store data first on device and then transmit them to server at predefined intervals or once connected to Wi-Fi
    - e.g., for smartphones, if no Wi-Fi connection available for longer time, more expensive cell connection is used

- Other systems require researcher to collect device and download data manually
    - e.g., research-grade accelerometers

Storage

- Size of data = Sampling rate * Field period
  - Even small samples, might produce *"Big Data"*
  - e.g., accelerometer data (three coordinates) collected at 60 Hz (i.e., 60 measurements per sec) for 10 min creates 108,000 data points per participant
- ...think whether you really need this amount of data and if so, have the appropriate infrastructure ready
- Processing data on device and only transmitting processed/aggregated data saves storage for researchers
  - e.g., using Google API to automatically classify accelerometer data into transportation modes
  - e.g., OCR for receipt scanning

Costs

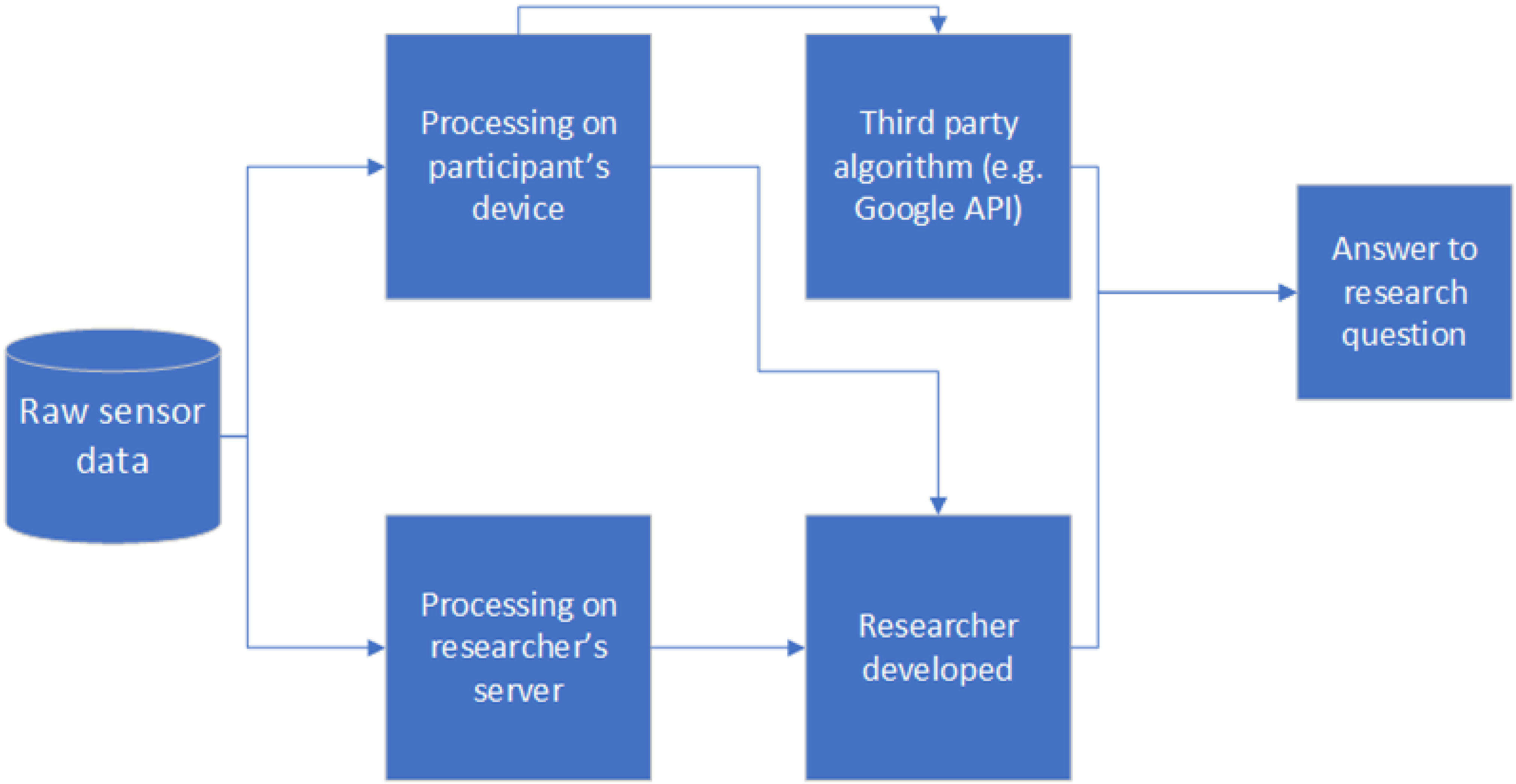
- Sensor measurement might seem relatively inexpensive because of small/no marginal costs of additional data
- Costs come from various sources
  - App development & maintenance (potentially for multiple OS)
  - Loaner devices
  - Incentives for participants
  - Technical support (e.g., hotline)
  - Storage infrastructure
  - Data handling & analytical skills (e.g., data wrangling, working with machine learning algorithms, data linkage, etc.)
  - Data visualization skills (e.g., GIS expertise)

Data from sensors, apps, and wearables

Data from wearables, apps, & sensors

- From raw data to insights
- Examples of (raw) data
- Processing raw data

From raw data to insights



Examples of sensor data: Accelerometer


```
SMotion_Time_Stamp;SMotion_without_Gravity;SM_with_Gravity
323,337,354,370,386,404,421,437,454,471,487,504,520,538,564,570,587,604,721,737,758,770,786,804,821,837,854,871,887,904,920,937,954,970,987,1004,1021,1
,3954,3971,3987,4004,4021,4037,4054,4071,4087,4104,4121,4137,4154,4171,4187,4204,4221,4238,4254,4271,4287,4304,4321,4337,4354,4371,4387,4404,4421,4437,
2,0.52,0.30,0.12,0.18,0.12,0.10,0.10,0.17,0.34,0.29,0.22,0.16,0.18,0.26,0.19,0.23,0.16,0.17,0.13,0.11,0.05,0.05,0.11,0.13,0.10,0.10,0.37,0.12,0.07,0.08,
06,0.08,0.12,0.22,0.22,0.13,0.37,0.81,0.54,0.28,0.29,0.44,0.28,0.23,0.26,0.32,0.23,0.32,0.36,0.18,0.19,0.24,0.31,0.20,0.13,0.21,0.23,0.21,0.16,0.13,0.1
.15,9.80,9.88,10.04,10.47,9.85,10.02,10.02,10.12,10.08,9.61,9.58,9.81,10.12,10.08,9.96,9.89,9.43,9.37,9.51,9.98,10.07,10.00,9.83,9.84,9.76,9.71,9.83,9.
.85,9.84,9.87,9.91,9.90,9.89,9.80,9.75,9.75,9.75,9.89,9.94,9.95,9.95,9.84,9.95,10.07,9.92,9.97,9.71,9.76,9.61,9.53,9.55,9.44,9.41,9.62,9.78,10.00,10.10,
222,237,254,274,288,304,321,337,354,372,388,404,422,438,455,472,488,504,522,538,554,572,588,604,621,638,673,688,710,721,738,754,772,791,805,821,838,854,
3838,3854,3871,3887,3904,3921,3938,3954,3971,3988,4006,4021,4038,4054,4071,4088,4104,4121,4138,4155,4172,4188,4204,4222,4238,4255,4271,4288,4304,4322,4
,0.08,0.07,0.10,0.11,0.10,0.15,0.18,0.12,0.19,0.07,0.15,0.08,0.14,0.30,0.16,0.12,0.18,0.10,0.13,0.06,0.04,0.04,0.12,0.35,0.48,0.28,0.31,0.31,0.18,0.11,
5,0.53,0.14,0.23,0.16,0.15,0.19,0.25,0.22,0.23,0.09,0.26,0.22,0.19,0.12,0.10,0.13,0.11,0.12,0.12,0.19,0.17,0.08,0.06,0.08,0.08,0.02,0.01,0.04,0.04,0.11,
.91,9.78,9.82,9.82,9.90,9.86,9.76,9.82,9.82,9.90,9.88,9.86,9.93,9.93,9.87,9.82,9.67,9.77,9.84,10.00,10.02,9.90,9.84,9.77,9.76,9.77,9.83,9.85,9.97,10.01,
111,128,144,452,569,587,619,637,717,734,751,769,786,802,819,836,853,871,888,906,923,940,958,974,990,1008,1025,1042,1059,1076,1093,1110,1127,1144,1161,1
,4254,4271,4288,4306,4324,4340,4358,4374,4391,4408,4425,4442,4459,4476,4493,4511,4529,4547,4565,4582,4599,4616,4633,4651,4668,4684,4702,4719,4736,4753,
6,0.25,0.15,0.15,0.37,0.29,0.13,0.30,0.29,0.24,0.46,0.52,0.42,0.41,0.18,0.10,0.26,0.31,0.29,0.36,0.43,0.31,0.15,0.15,0.14,0.13,0.23,0.35,0.34,0.30,0.31,
36,0.25,0.57,0.36,0.37,0.36,0.31,0.36,0.30,0.26,0.20,0.20,0.10,0.10,0.09,0.19,0.18,0.11,0.38,0.23,0.23,0.51,0.56,0.01,0.60,0.68,0.66,0.35,0.24,0.43,0.2
9.67,9.59,9.57,9.53,9.58,9.68,9.68,9.70,9.67,9.59,9.62,9.58,9.65,9.56,9.46,9.40,9.42,9.42,9.52,9.64,9.63,9.67,9.67,9.68,9.77,9.87,9.70,9.70,9.51,9.32,9.
,9.56,9.58,9.58,9.55,9.53,9.59,9.63,9.65,9.61,9.60,9.60,9.54,9.53,9.57,9.58,9.58,9.57,9.56,9.59,9.68,9.63,9.63,9.63,9.67,9.66,9.60,9.60,9.57,9.55,9.54,
183,199,216,236,249,266,282,299,316,333,348,366,382,399,416,433,449,465,482,500,515,532,549,566,582,599,616,632,649,666,682,699,716,733,749,766,783,799,
,0.07,0.09,0.10,0.11,0.08,0.09,0.07,0.07,0.07,0.06,0.04,0.07,0.12,0.11,0.06,0.10,0.15,0.14,0.21,0.08,0.08,0.08,0.13,0.50,0.38,0.29,0.34,0.36,0.47,0.46,
,9.93,9.85,9.65,9.53,9.44,10.19,10.04,9.91
137,608,725,847,852,852,853,855,856,856,860,863,863,864,865,866,866,866,867,867,867,868,870,917,918,918,926,927,927,947,989,990,990,991,991,992,1001,10
2797,2807,2821,2831,2840,2855,2865,2875,2896,2906,2927,2936,2946,2956,2966,2976,2997,3007,3017,3027,3052,3069,3077,3090,3105,3115,3125,3135,3146,3155,3
,5192,5202,5213,5222,5232,5243,5252,5263,5273,5293,5304,5314,5324,5335,5345,5366,5376,5386,5406,5417,5429,5480,5480,5481,5482,5482,5532,5533,5533,5533,
1,7451,7485,7486,7486,7511,7511,7512,7518,7529,7538,7549,7559,7569,7579,7589,7600,7610,7621,7631,7641,7651,7661,7671,7681,7691,7701,7711,7722,7731,7741,
61,9861,9862,9864,9865,9865,9865,9866,9866,9866,9867,9867,9867,9868,9868,9868,9869,9869,9869,9922,9922,9922,9922,9923,9927,9964,9965,9968,9978,10022,10
,11612,11621,11632,11641,11651,11661,11671,11681,11691,11701,11712,11722,11732,11742,11752,11762,11772,11782,11793,11803,11814,11827,11837,11847,11857,
06,13516,13537,13547,13556,13569,13578,13590,13611,13641,13665,13681,13691,13722,13732,13745,13775,13784,13795,13805,13836,13846,13856,13866,13877,1388
5821,15832,15842,15859,15863,15875,15883,15896,15913,15922,15932,15943,15953,15963,15973,15983,15993,16004,16015,16025,16037,16047,16056,16067,16078,16
```

Examples of sensor data: App use

```
PackageName,TimestampOnStart,TimeInfoOnStart_TimeSource,TimestampOnEnd,TimeInfoOnEnd_TimeSource,AppUsageTime,SessionTotalTxBytes,SessionTotalRxBytes,AppName
com.android.settings,2018-06-06 12:03:48.674 +0200,Device,2018-07-03 12:03:54.717 +0200,Device,5890,1599939,603658,Unknown,Unknown
com.p3group.insight.iab.smart,2018-06-18 14:17:23.084 +0200,Device,2018-06-18 14:17:42.491 +0200,Device,19405,0,0,Unknown,Unknown
com.sec.android.app.launcher,2018-06-18 14:17:42.491 +0200,Device,2018-06-18 14:17:45.403 +0200,Device,2304,0,0,Unknown,Unknown
com.android.chrome,2018-06-18 15:47:01.564 +0200,NTP,2018-06-18 15:50:07.324 +0200,NTP,185703,399321,3037467,Chrome Browser - Google,Communication
com.sec.android.app.launcher,2018-06-18 16:59:40.785 +0200,NTP,2018-06-18 16:59:50.649 +0200,NTP,9862,0,0,Unknown,Unknown
org.telegram.messenger,2018-06-18 19:04:53.836 +0200,NTP,2018-06-18 19:05:05.701 +0200,NTP,11864,0,0,Telegram,Communication
com.sec.android.app.launcher,2018-06-18 19:36:35.838 +0200,NTP,2018-06-18 19:36:38.682 +0200,NTP,2844,0,0,Unknown,Unknown
org.telegram.messenger,2018-06-18 19:36:38.682 +0200,NTP,2018-06-18 19:36:55.756 +0200,NTP,16761,0,0,Telegram,Communication
com.android.chrome,2018-06-18 19:36:55.756 +0200,NTP,2018-06-18 19:40:33.864 +0200,NTP,217612,66950,1225595,Chrome Browser - Google,Communication
org.telegram.messenger,2018-06-18 19:40:33.864 +0200,NTP,2018-06-18 19:48:10.711 +0200,NTP,456189,10988,6702,Telegram,Communication
org.telegram.messenger,2018-06-18 20:06:34.802 +0200,NTP,2018-06-18 20:07:13.155 +0200,NTP,38352,639,579,Telegram,Communication
de.hafas.android.db,2018-06-18 22:55:42.359 +0200,NTP,2018-06-18 22:57:04.772 +0200,NTP,82376,104245,714228,DB Navigator,Maps & Navigation
com.sec.android.app.clockpackage,2018-06-19 06:00:05.904 +0200,NTP,2018-06-19 06:00:40.337 +0200,NTP,34406,0,0,Unknown,Unknown
com.sec.android.app.launcher,2018-06-19 06:00:40.337 +0200,NTP,2018-06-19 06:00:46.010 +0200,NTP,5148,0,0,Unknown,Unknown
com.sec.android.app.clockpackage,2018-06-19 06:00:46.010 +0200,NTP,2018-06-19 06:00:55.898 +0200,NTP,9599,0,0,Unknown,Unknown
com.sec.android.app.clockpackage,2018-06-19 06:15:52.878 +0200,NTP,2018-06-19 06:16:10.146 +0200,NTP,17268,0,0,Unknown,Unknown
com.sec.android.app.clockpackage,2018-06-19 08:20:16.361 +0200,NTP,2018-06-19 08:20:19.305 +0200,NTP,2945,0,0,Unknown,Unknown
com.google.android.music,2018-06-19 08:20:21.202 +0200,NTP,2018-06-19 08:21:18.689 +0200,NTP,57420,35279,63680,Google Play Music,Music & Audio
com.whatsapp,2018-06-19 14:59:07.227 +0200,NTP,2018-06-19 14:59:36.351 +0200,NTP,29076,252113,11862,WhatsApp Messenger,Communication
com.sec.android.app.launcher,2018-06-19 17:02:20.977 +0200,NTP,2018-06-19 17:02:29.893 +0200,NTP,8710,0,0,Unknown,Unknown
com.google.android.apps.maps,2018-06-19 17:02:52.774 +0200,NTP,2018-06-19 17:04:01.406 +0200,GPS,69035,117029,882507,Maps - Navigation & Transit,Travel & Lo
com.google.android.apps.maps,2018-06-19 17:30:26.277 +0200,GPS,2018-06-19 17:30:34.221 +0200,GPS,7941,1156,759,Maps - Navigation & Transit,Travel & Local
org.telegram.messenger,2018-06-19 19:48:37.730 +0200,NTP,2018-06-19 19:48:43.653 +0200,NTP,5920,362,245,Telegram,Communication
org.telegram.messenger,2018-06-19 20:17:42.599 +0200,NTP,2018-06-19 20:18:12.791 +0200,NTP,30191,2952,2446,Telegram,Communication
com.sec.android.app.launcher,2018-06-19 20:22:42.508 +0200,NTP,2018-06-19 20:22:46.876 +0200,NTP,4368,0,0,Unknown,Unknown
com.samsung.android.email.provider,2018-06-19 20:22:46.876 +0200,NTP,2018-06-19 20:22:53.042 +0200,NTP,5547,1310,5078,Unknown,Unknown
com.sec.android.app.launcher,2018-06-19 20:22:53.042 +0200,NTP,2018-06-19 20:22:56.010 +0200,NTP,2667,0,0,Unknown,Unknown
com.android.chrome,2018-06-19 20:22:56.010 +0200,NTP,2018-06-19 20:23:59.152 +0200,NTP,61363,300683,2813090,Chrome Browser - Google,Communication
com.android.chrome,2018-06-19 20:24:31.686 +0200,NTP,2018-06-19 20:26:31.206 +0200,NTP,119514,148480,1220440,Chrome Browser - Google,Communication
com.whatsapp,2018-06-19 22:36:01.409 +0200,NTP,2018-06-19 22:36:07.606 +0200,NTP,6142,323,356,WhatsApp Messenger,Communication
com.sec.android.app.clockpackage,2018-06-20 06:00:03.358 +0200,NTP,2018-06-20 06:05:00.330 +0200,NTP,296954,0,0,Unknown,Unknown
com.sec.android.app.clockpackage,2018-06-20 06:05:02.349 +0200,NTP,2018-06-20 06:06:02.618 +0200,NTP,60235,0,0,Unknown,Unknown
com.whatsapp,2018-06-20 06:06:02.618 +0200,NTP,2018-06-20 06:06:11.920 +0200,NTP,8939,0,0,WhatsApp Messenger,Communication
com.sec.android.app.clockpackage,2018-06-20 06:10:04.246 +0200,NTP,2018-06-20 06:11:01.198 +0200,NTP,56951,0,0,Unknown,Unknown
```

Examples of sensor data: Apple Health


```
<!--  HealthKit Export Version: 8  -->
▼<!--
   Note: Any Records that appear as children of a correlation also appear as top-level records in this document.
-->
▼<!--
   Note: Heart Rate Variability records captured by Apple Watch may include an associated list of instantaneous beats-per-minute readings.
-->
▼<HealthData locale="en_AT">
  <ExportDate value="2019-06-18 10:27:20 +0200"/>
  <Me HKCharacteristicTypeIdentifierDateOfBirth="" HKCharacteristicTypeIdentifierBiologicalSex="HKBiologicalSexNotSet"
  HKCharacteristicTypeIdentifierBloodType="HKBloodTypeNotSet" HKCharacteristicTypeIdentifierFitzpatrickSkinType="HKFitzpatrickSkinTypeNotSet"/>
  <Record type="HKQuantityTypeIdentifierStepCount" sourceName="Florian's iPhone (2)" sourceVersion="11.4" device="<<HKDevice: 0x280f8e170>, name:iPhone, manufacturer:Apple,
  model:iPhone, hardware:iPhone10,6, software:11.4>" unit="count" creationDate="2018-07-05 19:05:02 +0200" startDate="2018-07-05 18:26:11 +0200" endDate="2018-07-05 18:36:10
  +0200" value="88"/>
  <Record type="HKQuantityTypeIdentifierStepCount" sourceName="Florian's iPhone (2)" sourceVersion="11.4" device="<<HKDevice: 0x280f8dd60>, name:iPhone, manufacturer:Apple,
  model:iPhone, hardware:iPhone10,6, software:11.4>" unit="count" creationDate="2018-07-05 19:05:02 +0200" startDate="2018-07-05 18:36:10 +0200" endDate="2018-07-05 18:42:08
  +0200" value="202"/>
  <Record type="HKQuantityTypeIdentifierStepCount" sourceName="Florian's iPhone (2)" sourceVersion="11.4" device="<<HKDevice: 0x280f8d450>, name:iPhone, manufacturer:Apple,
  model:iPhone, hardware:iPhone10,6, software:11.4>" unit="count" creationDate="2018-07-05 19:05:02 +0200" startDate="2018-07-05 18:42:08 +0200" endDate="2018-07-05 18:51:09
  +0200" value="87"/>
  <Record type="HKQuantityTypeIdentifierStepCount" sourceName="Florian's iPhone (2)" sourceVersion="11.4" device="<<HKDevice: 0x280f8c280>, name:iPhone, manufacturer:Apple,
  model:iPhone, hardware:iPhone10,6, software:11.4>" unit="count" creationDate="2018-07-05 19:29:31 +0200" startDate="2018-07-05 18:58:47 +0200" endDate="2018-07-05 19:07:18
  +0200" value="53"/>
  <Record type="HKQuantityTypeIdentifierStepCount" sourceName="Florian's iPhone (2)" sourceVersion="11.4" device="<<HKDevice: 0x280f8c230>, name:iPhone, manufacturer:Apple,
  model:iPhone, hardware:iPhone10,6, software:11.4>" unit="count" creationDate="2018-07-05 19:29:31 +0200" startDate="2018-07-05 19:07:18 +0200" endDate="2018-07-05 19:15:14
  +0200" value="52"/>
  <Record type="HKQuantityTypeIdentifierStepCount" sourceName="Florian's iPhone (2)" sourceVersion="11.4" device="<<HKDevice: 0x280f8c8c0>, name:iPhone, manufacturer:Apple,
  model:iPhone, hardware:iPhone10,6, software:11.4>" unit="count" creationDate="2018-07-05 19:29:31 +0200" startDate="2018-07-05 19:15:14 +0200" endDate="2018-07-05 19:25:05
  +0200" value="202"/>
  <Record type="HKQuantityTypeIdentifierStepCount" sourceName="Florian's iPhone (2)" sourceVersion="11.4" device="<<HKDevice: 0x280f8cb40>, name:iPhone, manufacturer:Apple,
  model:iPhone, hardware:iPhone10,6, software:11.4>" unit="count" creationDate="2018-07-05 20:04:37 +0200" startDate="2018-07-05 19:31:27 +0200" endDate="2018-07-05 19:40:01
  +0200" value="137"/>
  <Record type="HKQuantityTypeIdentifierStepCount" sourceName="Florian's iPhone (2)" sourceVersion="11.4" device="<<HKDevice: 0x280f8cdc0>, name:iPhone, manufacturer:Apple,
  model:iPhone, hardware:iPhone10,6, software:11.4>" unit="count" creationDate="2018-07-05 20:04:37 +0200" startDate="2018-07-05 19:41:27 +0200" endDate="2018-07-05 19:49:17
  +0200" value="8"/>
```

Examples of sensor data: GPS and trip motives

	device_id	latitude	longitude	accuracy	speed	altitude	timestamp
1:	23	52.09460	5.134593	15.204	0	54.7	2018-10-31 22:53:22
2:	23	52.09460	5.134593	15.204	0	54.7	2018-10-31 22:54:22
3:	23	52.09460	5.134593	15.204	0	54.7	2018-10-31 22:56:40
4:	23	52.09460	5.134593	15.204	0	54.7	2018-10-31 22:57:40
5:	23	52.09460	5.134593	15.204	0	54.7	2018-10-31 22:59:05

38524:	23	52.09464	5.134572	15.175	0	54.7	2018-11-12 00:00:33
38525:	23	52.09464	5.134572	15.175	0	54.7	2018-11-12 00:00:33

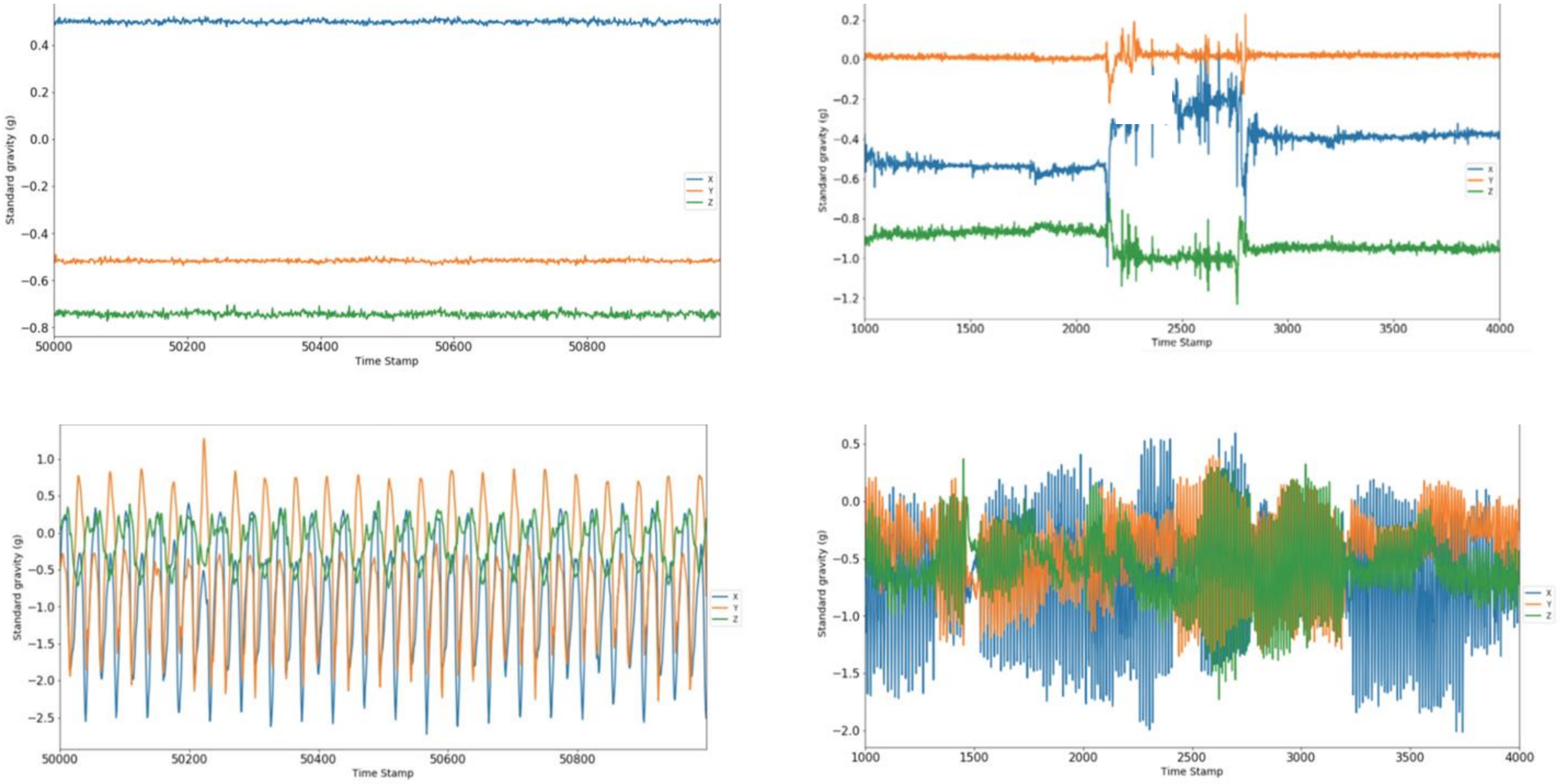
	device_id	local_stop_id	begin_timestamp	end_timestamp
1:	23	7	2018-10-31 17:05:47	2018-10-31 17:11:51
2:	23	11	2018-10-31 17:26:56	2018-10-31 17:31:39
3:	23	5	2018-10-31 17:32:51	2018-10-31 17:40:09
4:	23	4	2018-10-31 17:45:13	2018-10-31 19:03:58
5:	23	8	2018-10-31 19:04:08	2018-11-01 12:53:08
6:	23	9	2018-11-01 13:00:52	2018-11-01 15:47:21

	device_id	local_stop_visit_id	motive
1:	23	3	Home
2:	23	2	PaidWork
3:	23	1	Home
4:	23	7	Transfer
5:	23	6	Transfer
6:	23	4	Home

Processing raw data

- Data needs to be cleaned and processed before analysis (*Data wrangling/munging*)
  - This usually takes much longer than data analysis (80/20 rule)
- Aggregation of raw data to meaningful data point level
  - What is “meaningful” depends on research and use of data
- Processing of raw data can happen on
  - User’s device using (built-in) third party or researcher-developed algorithm
    - Preserves storage and protects privacy
    - No access to raw data
  - Researcher’s server
    - Full control over data processing
    - All data needs to be transfered

Exercise: Detect types of activity



Mulder et al. (2019)

Model building pipeline

Data Cleaning & pre-processing

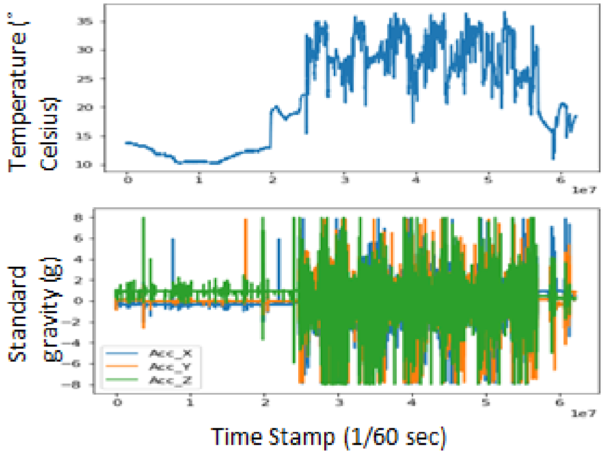
- Removal non-wear time
- Removal of high frequency (frequency higher than 15 Hz)
- Data with wear time less than 7 days discarded

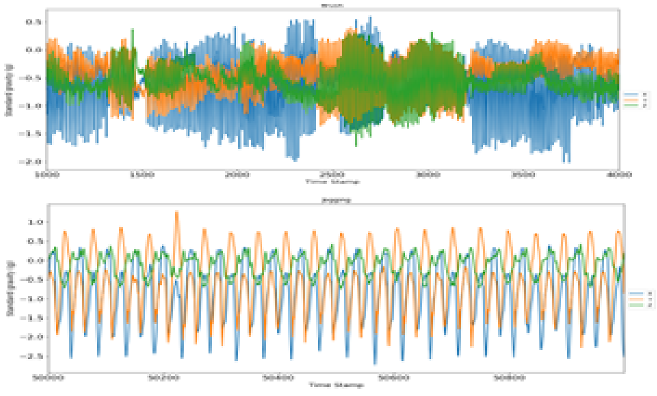
Feature Engineering

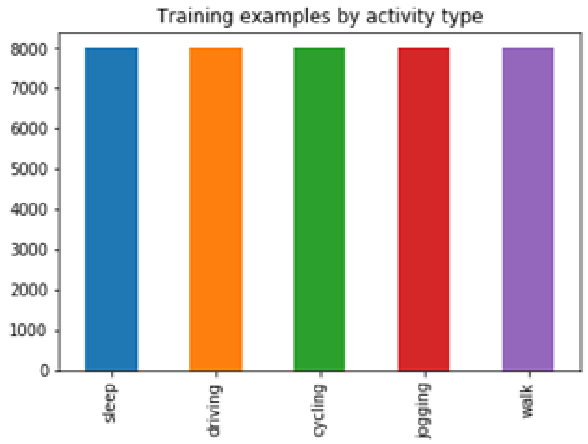
- **Time domain:** X, Y, Z, temperature, mean, median, standard deviation, RMS, percentile distribution
- **Frequency domain:** FT, dominant frequency selection, power of signal

Model building & validation

- Optimizing the epoch time
- Preparing balanced dataset
- Train/test splitting of 80%/20%
- Training and validation of the model (SVM, RF, and LR model)



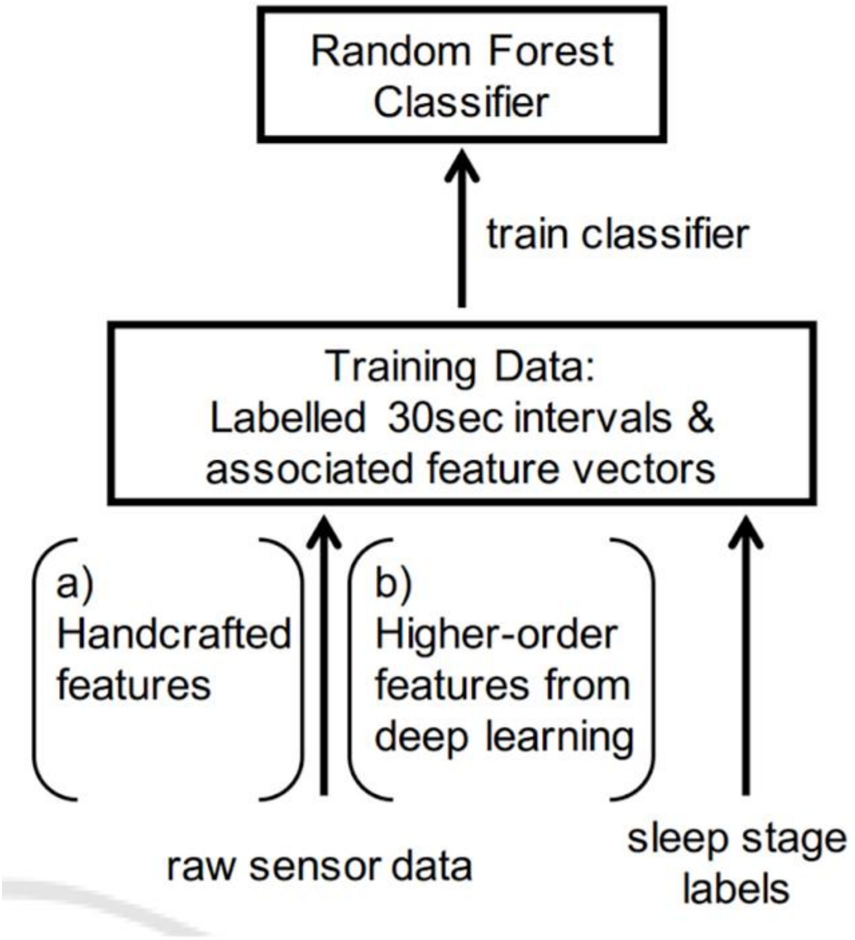




Mulder et al. (2019)

Example: Sleep detection

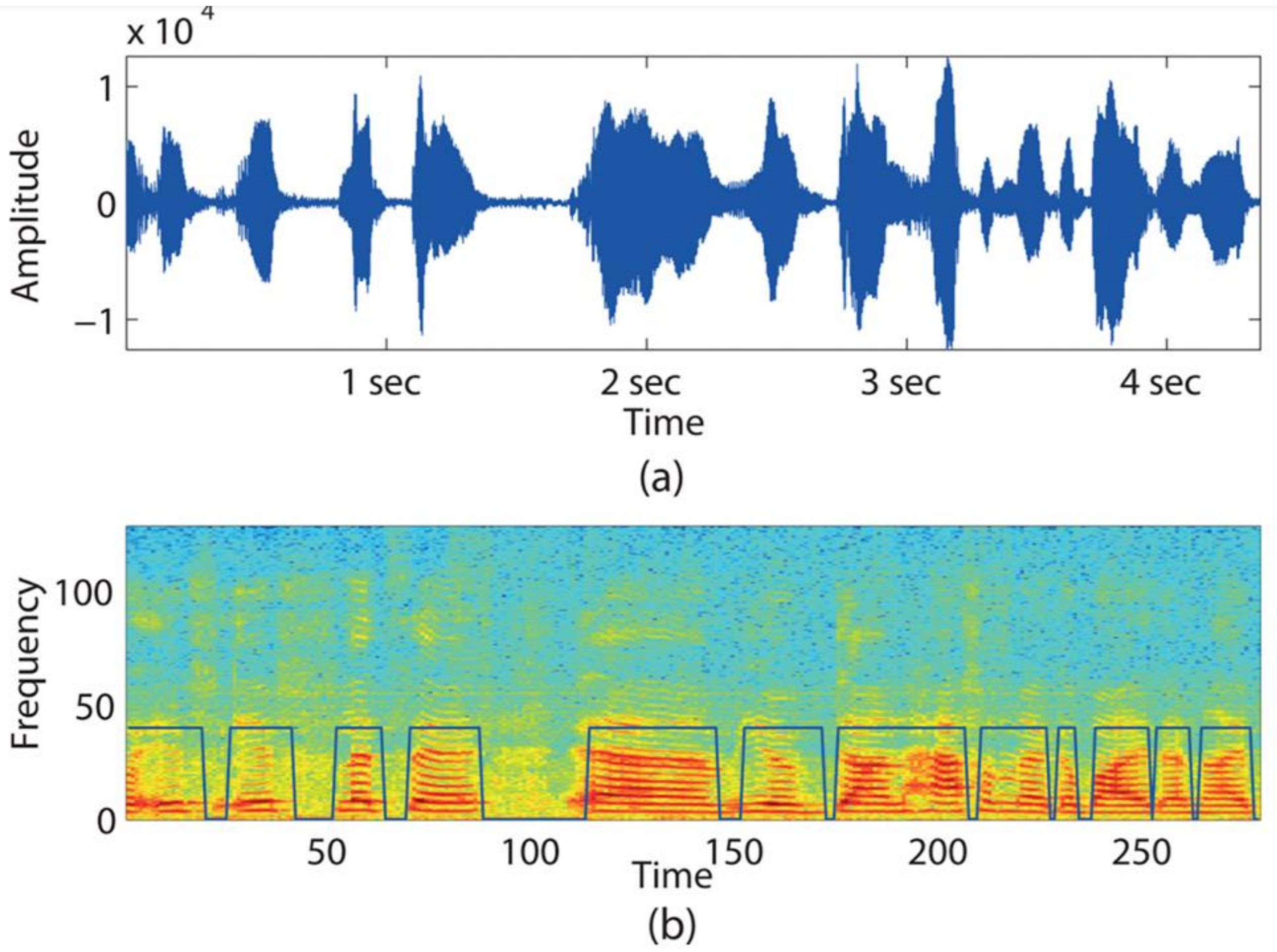
- Identifying sleep stages from heart rate and
  movements from fitness bracelet
- Identifying sleep duration using smartphone features:
  - e.g., light, microphone, usage (phone lock, recharging, phone off), accelerometer ("Best effort sleep model")



Reimer et al. (2017)

Example: Detecting conversations

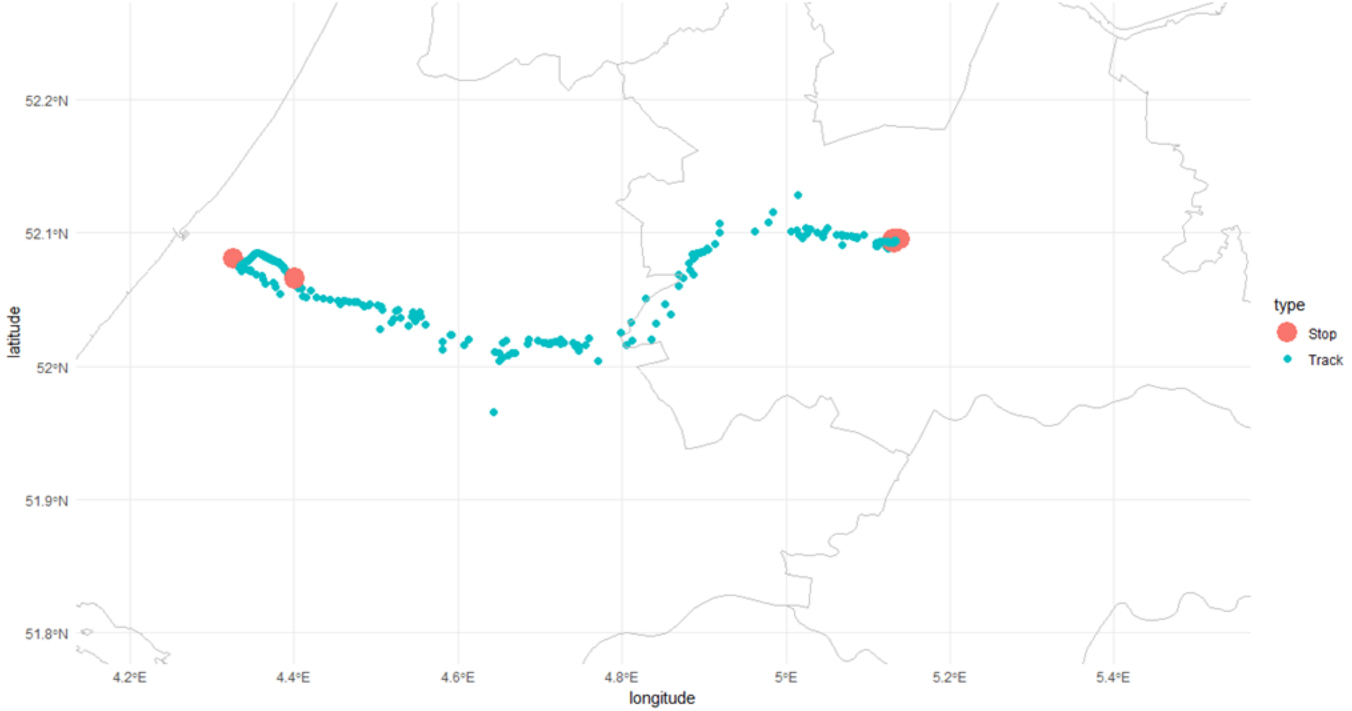
- Using smartphone microphone to detect personal conversations
  - Microphone always on but content of conversation not transmitted
  - Outcome of inference: 0 = no conversation, 1 = conversation
- Processing raw data on device
  - Privacy sensitive classifiers (Wyatt et al. 2007)
  - Transferred data only includes aggregated information

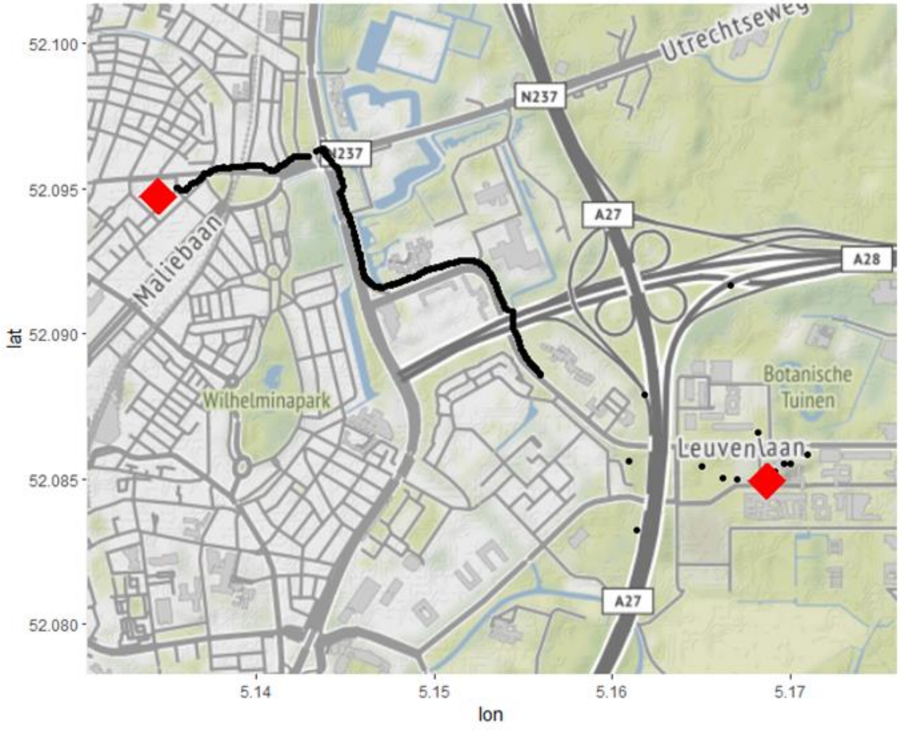


Rabbi et al. (2011)

Example: GPS tracks and stop detection (McCool et al. 2019)

- Stops defined based on “static” location: radius has to be (pre)defined by researcher





Errors when collecting, processing, and interpreting sensor data

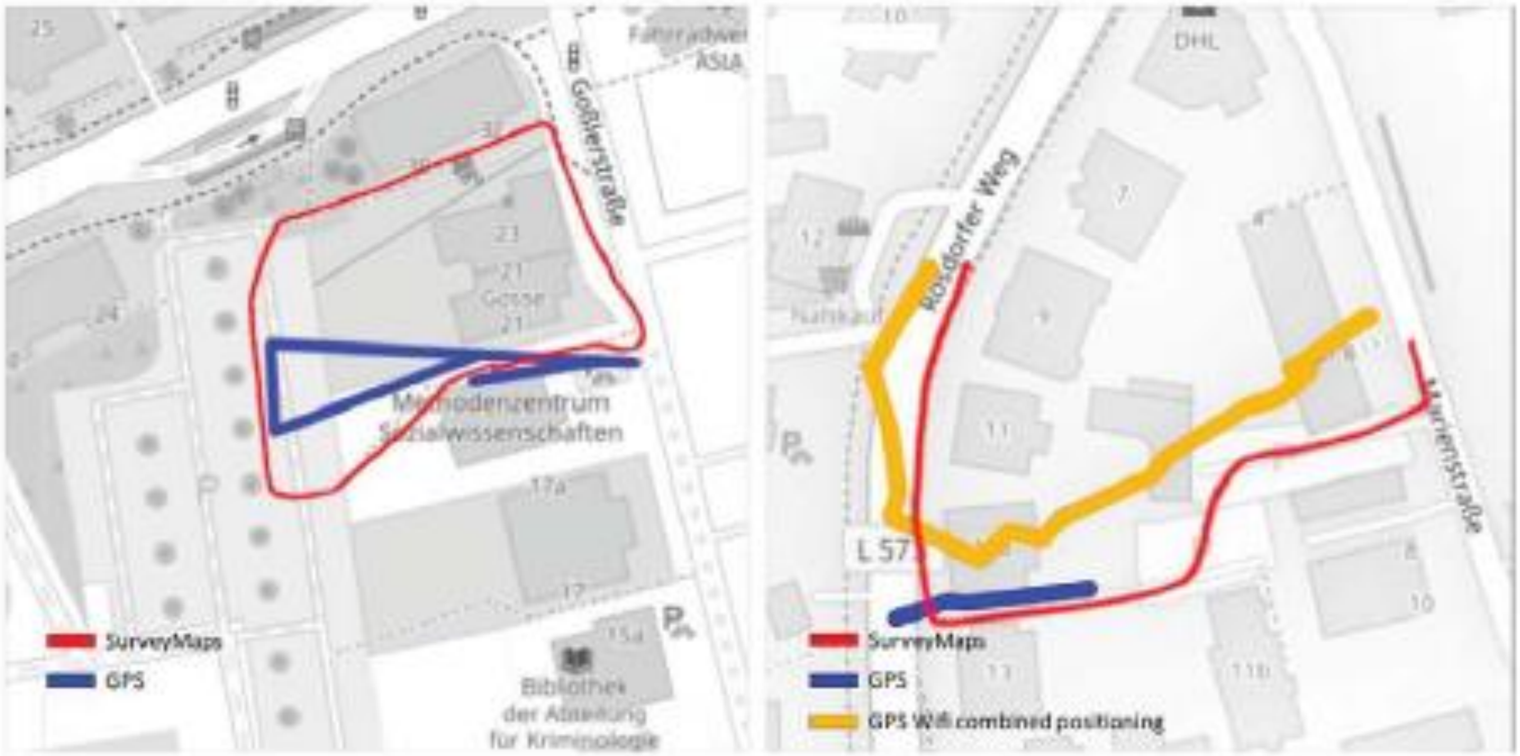
Errors when collecting, processing, and interpreting sensor data

- Tempting to assume that by removing human cognition and social interaction from “passive” sensor data collection, we eliminate all measurement error

- But errors might still arise when…
  - ...collecting the data
  - ...processing the data
  - ...interpreting the data

Errors during data collection

- Sensor-based errors/differences
  - Differences between types of sensors as well as brands and models of devices
  - Not one sensor/device per se better than others, depends on what should be measured under what circumstances



Schlosser et al. (2019)

Exercise

You want to measure *mobility*, *physical activity*, and *sleep*. How can the way people handle their smartphones introduce error?

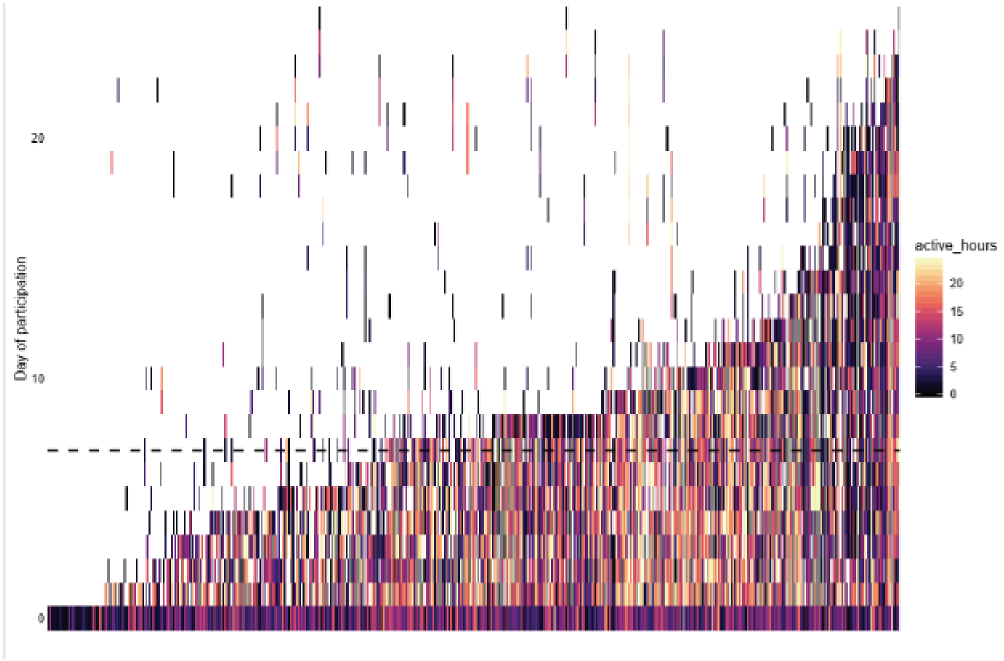
Errors during data collection

- Sensor-based errors/differences
- Device handling
    - Measurement might differ depending on where/ how sensor/device is worn
        - e.g., differences in how men and women carry around smartphones

Sztyler et al. (2017)

Errors during data collection

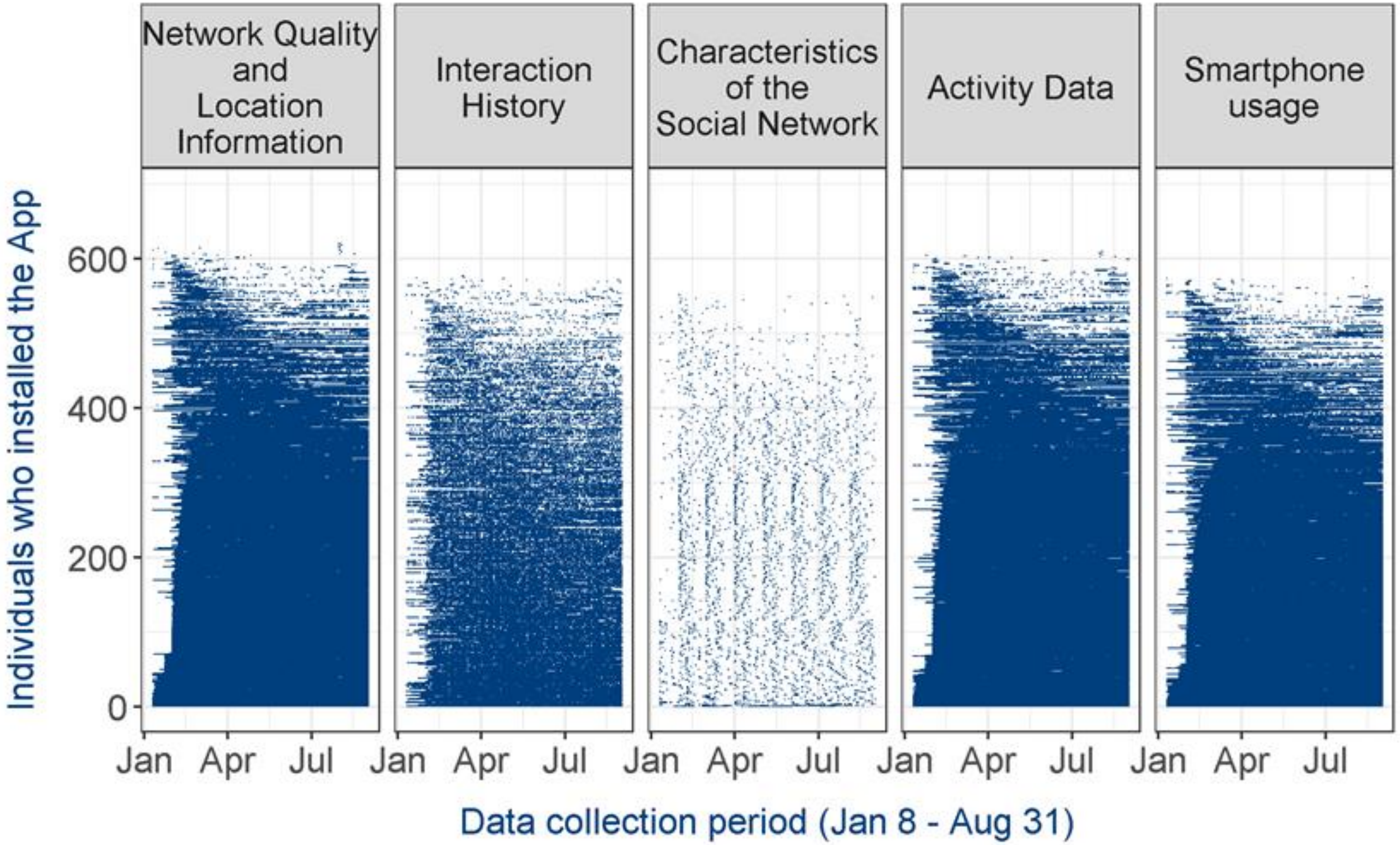
- Sensor-based errors
- Device handling
- Missing data
  - Technical issues:
    - Urban canyons, underground, etc. when collecting GPS
    - Device out of power or sleep mode
    - iOS blocks collection of location in background
    - ...
  - Noncompliance:
    - Leaving device at home
    - Deliberately turning device off at certain locations or times
    - Forgetting to turn device back on again
    - Missing permissions
    - ...



Active hours per user

McCool et al. (2019)

Example: Missing data (Bähr et al. 2022)



Errors during data collection

- Sensor-based errors
- Device handling
- Missing data
- Erroneous/Invalid data
  - e.g., fake GPS apps, VPN

Source: Sebastian Bähr

Errors during data collection

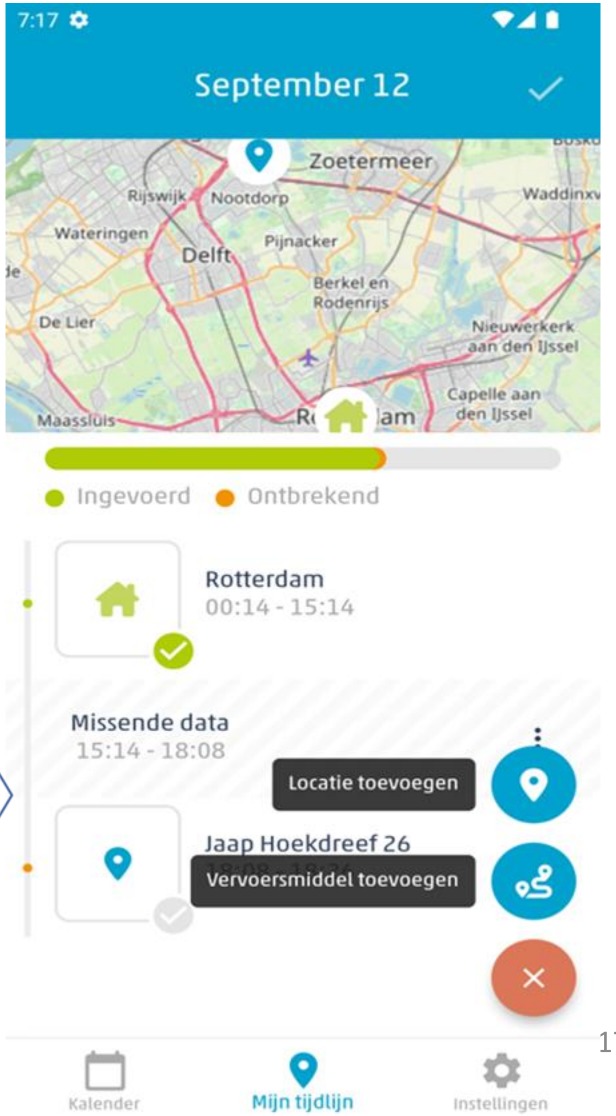
- Sensor-based errors
- Device handling
- Missing data
- Erroneous data
- Providing feedback & measurement reactivity
    - e.g., participants show 7% more physical activity when wearing Fitbit (with feedback) compared to when wearing GENEActive (no feedback) (Darling et al. 2021)

Source: https://twitter.com/mbrennanchina/status/1128201958962032641

Providing feedback to participants

Household Budget Survey (HBS)
fall 2021,
NL, ES, LU
N=3916,
Completion = 16%
No influence of feedback on representativeness, data quality

Travel app possibility to provide context to passive data, add data

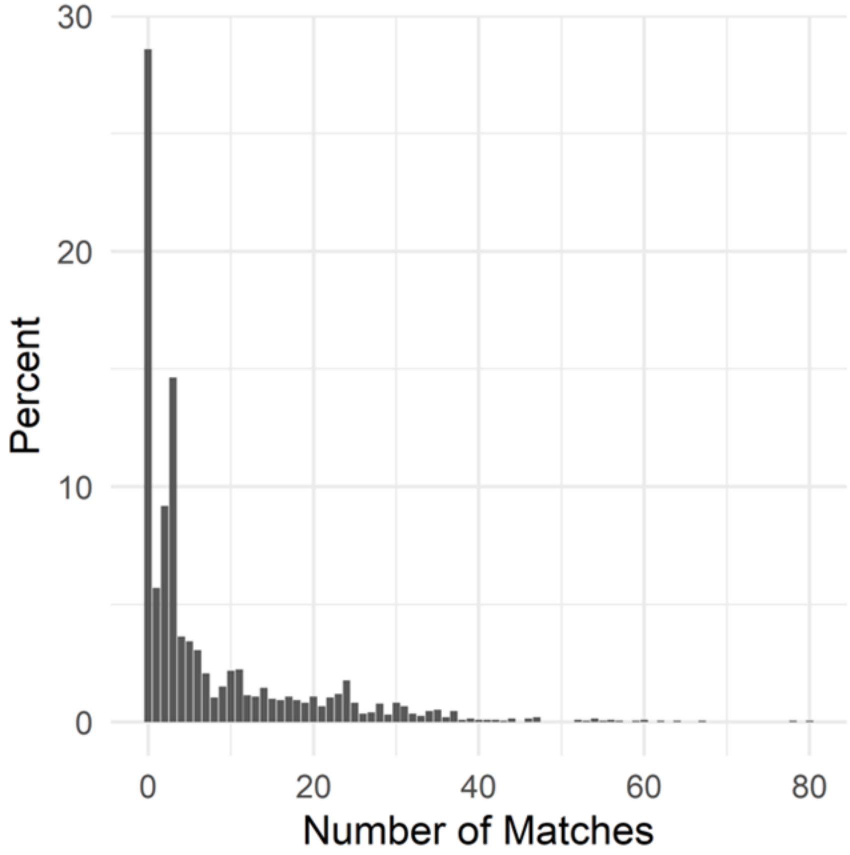


Errors during inference & interpretation

- Raw sensor data must be processed and classified to infer behavior
- "Black box" approach when using third-party algorithm to classify data on device
  - What looks like raw data to researcher is actually (heavily) pre-processed
  - e.g., activity classification was trained based on data from young adults ("WEIRDOS" ©Mick P. Couper) → used to classify behavior of general population
  - e.g., smartphone forgotten at home in a bag → respondent is asleep
- Self-report still needed for validation

Errors during inference & interpretation

- Errors can also arise when using third-party reference databases
- Challenges in matching places to points of interest (Eckman et al. 2020)
  - Data from multiple bases agreed only on 53 of 1,928 collected coordinates
  - Disagreement between data sources
    - “Strip mall problem”: grocery store on Google, liquor store on Foursquare, dentist’s office on Yelp
    - Agreement for very big stores, airports, etc.
  - Sometimes, too many matches
  - Sometimes, no matches at all



Eckman et al. (2020)

Additional resources

Selected resources for app development

- Commercial/Off-the-shelf existing platforms
  - Movisens: https://www.movisens.com/en/
  - MOTUS: https://www.motusresearch.io/en
  - Murmuras: https://murmuras.com/

- Commercial app builders (usually no special knowledge required)
  - Appypie https://www.appypie.com
  - Ethica Data: https://ethicadata.com/

Selected resources for app development

- App builders for specific OSs (require some programming knowledge)
  - Apple Research Kit: http://researchkit.org/
  - ResearchStack for Android: http://researchstack.org/

- Open source platforms/frameworks (require programming knowledge)
  - AWARE: https://awareframework.com/
  - Beiwe Research Platform: https://www.beiwe.org/
  - PACO: https://pacoapp.com/

Selected resources for EMA/ESM

- Specific EMA/ESM software
  - mEMA: https://ilumivu.com
  - ExpiWell: https://www.expiwell.com/
  - LifeData: https://www.lifedatacorp.com/ecological-momentary-assessment-app-2/
  - SEMA3: https://sema3.com/
  - Other online survey software, such as Blaise5 (https://blaise.com/products/blaise-5), can be used as sample management system that can send surveys at specific time
- Myin-Germeys, Inez, and Peter Kuppens. (Eds.). 2022. *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies.* (2nd ed.) Leuven: Center for Research on Experience Sampling and Ambulatory Methods Leuven

Other resources

- For visualization of location data:
  - Shiny app Utrecht University (R code): https://github.com/sobradob/shinyapp
- For data processing:
  - R package for log data analysis (Stachl): https://osf.io/ut42y/

Our book…

Keusch, Florian, Bella Struminskaya, Stephanie Eckman, and Heidi Guyer. forthcoming. *Data Collection with Wearables, Apps, and Sensors.* https://bookdown.org/wasbook_feedback/was/

Conferences

- Mobile Apps and Sensors in Surveys (MASS) Workshop
  Washington DC, March 2024
  https://massworkshop.org/

- General Online Research (GOR), Kassel September 2023 & Cologne February 2024
  http://gor.de

- ESRA 2025 @ Utrecht University