Data Collection with Apps, Sensors, and Wearables

Frauke Kreuter & Florian Keusch

AAPOR Webinar

November 14, 2019

@fraukolos
@floriankeusch

Acknowledgement

The material presented here is result of various research collaborations and joint teaching with:

Sebastian Bähr, Frederick Conrad, Mick Couper, Heidi Guyer, Georg-Christoph Haas, Jan Karem Höhne, Peter Lugtig, Bella Struminskaya, Mark Trappmann, and many more…

What we will talk about

- Why add apps, sensors, and wearables to surveys?
- What can we measure?
- What kind of research questions can be answered?
  - Case Study: IAB-SMART
- Study design considerations from the TSE perspective
- Processing data from sensors, apps, and wearables

Why add apps, sensors, and wearables to surveys?

Potential benefits of apps, sensors, & wearables

1. New forms of measurement
    - *In situ* measurement (e.g., Ecological Momentary Assessments - EMAs)
    - *Passive* measurement with sensors (e.g., automatic collection of location and activity)
    - Use of other device features for *active* measurement (e.g., photos, videos)

Potential benefits of apps, sensors, & wearables

1. New forms of measurement
2. More detailed data (frequency and intensity)
    - High frequency of measurement (e.g., accelerometer with 60Hz)
    - New types of information that cannot be self-reported (e.g., different stages of sleep)
    - Various data formats from one device (e.g., location, motion, pictures,…)

Potential benefits of apps, sensors, & wearables

1. New forms of measurement
2. More detailed data (frequency and intensity)
3. More accurate data (sometimes!)
    - No self-report = No recall error
    - No self-report = Potentially less social desirability

Potential benefits of apps, sensors, & wearables

1. New forms of measurement
2. More detailed data (frequency and intensity)
3. More accurate data (sometimes!)
4. Less response burden
    - Fewer survey questions have to be answered 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. New forms of measurement
2. More detailed data (frequency and intensity)
3. More accurate data (sometimes!)
4. Less response burden
5. Collecting data at scale
    - ~22,000 volunteer iPhone users downloaded *Mappiness* app and shared activities and affect (EMAs) plus geolocation (GPS) for 6 months (MacKarron & Murrato 2013)
    - >100,000 participants of the UK Biobank study wore wrist accelerometer for 7 days (Doherty et al. 2017)
    - 650 members of existing offline panel downloaded *IAB-SMART* app and responded to mini-surveys plus shared location, physical activity, smartphone use data for 6 months (Kreuter et al. 2018)

Potential benefits of apps, sensors, & wearables

1. New forms of measurement
2. More detailed data (frequency and intensity)
3. More accurate data (sometimes!)
4. Less response burden
5. Collecting data at scale
6. Introducing design to big data
    - Sensor data have many characteristics of Big Data (e.g., large volume, high velocity, variety of data formats)
    - Combining passive sensor data collection with self-reports through surveys introduces “design” to Big Data

Potential benefits of apps, sensors, & wearables

1. New forms of measurement
2. More detailed data (frequency and intensity)
3. More accurate data (sometimes!)
4. Less response burden
5. Collecting data at scale
6. Introducing design to big data
7. New research questions?

What can we measure?

Different devices with sensors

Source: https://www.youtube.com/watch?v=FEr9D2gIDXA

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/

Sensors in Samsung smartphones*

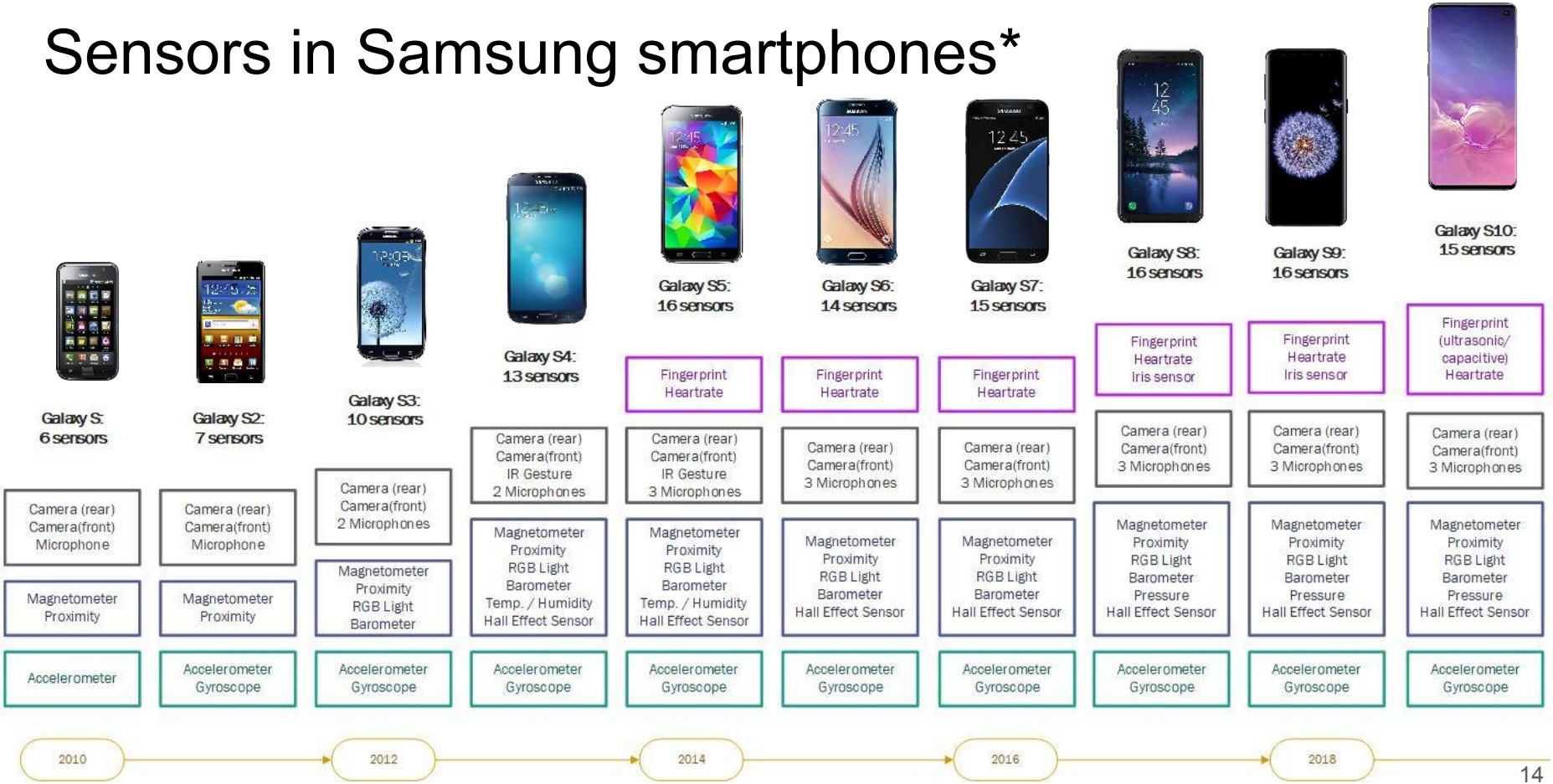
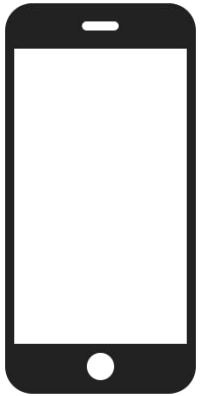


*Chart created by Bella Struminskaya

Native smartphone sensors



NFC

Heart rate sensor

Bluetooth

Air humidity sensor

Thermometer

Proximity sensor

Wi-Fi

Microphone

GPS

Light sensor

Cellular network

Camera

Fingerprint sensor

Compass

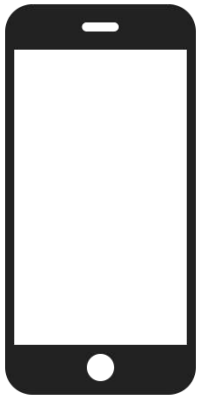
Barometer

Pedometer

Accelerometer

Gyroscope

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

Fingerprint sensor

Camera

Barometer

Compass

Accelerometer

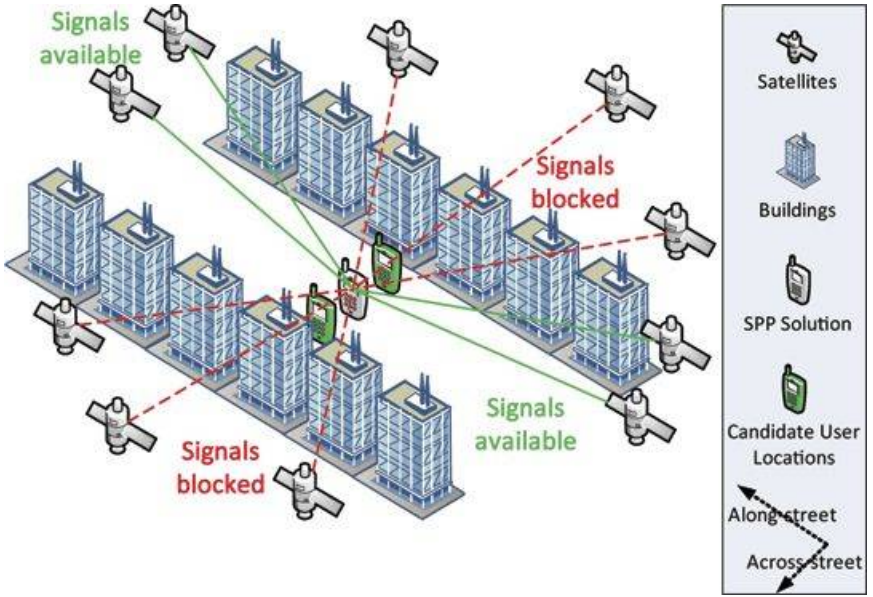
Pedometer

Gyroscope

Physical activity

Location sensors

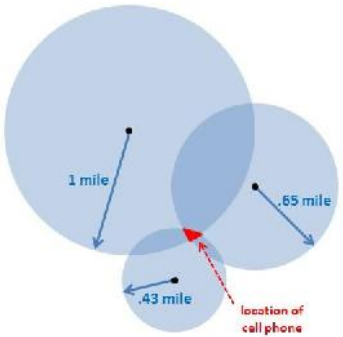
- GPS
  - Provides coordinates in longitude & latitude
    - Based on distance (= rate x time) to at least 4 satellites
  - Newest generation has accuracy within 30 centimeters
  - Works without cell/Internet connection
  - Performs worse in ‘urban canyons’, indoors, & underground
  - Constant GPS tracking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



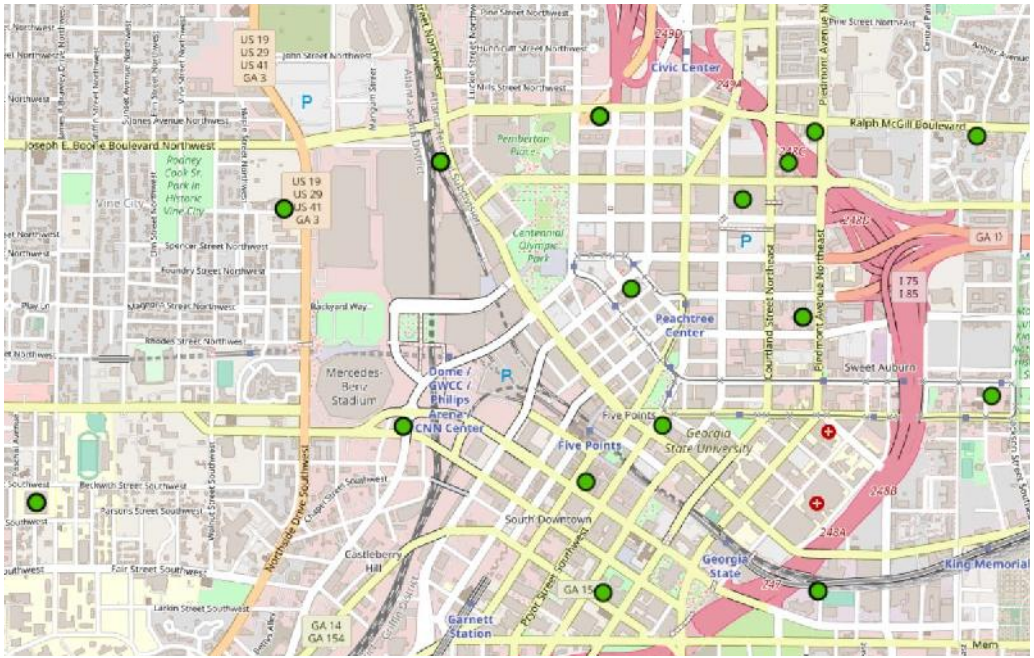
Picture source: https://www.gpsworld.com/wirelesspersonal-navigationshadow-matching-12550/

Location sensors

- 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



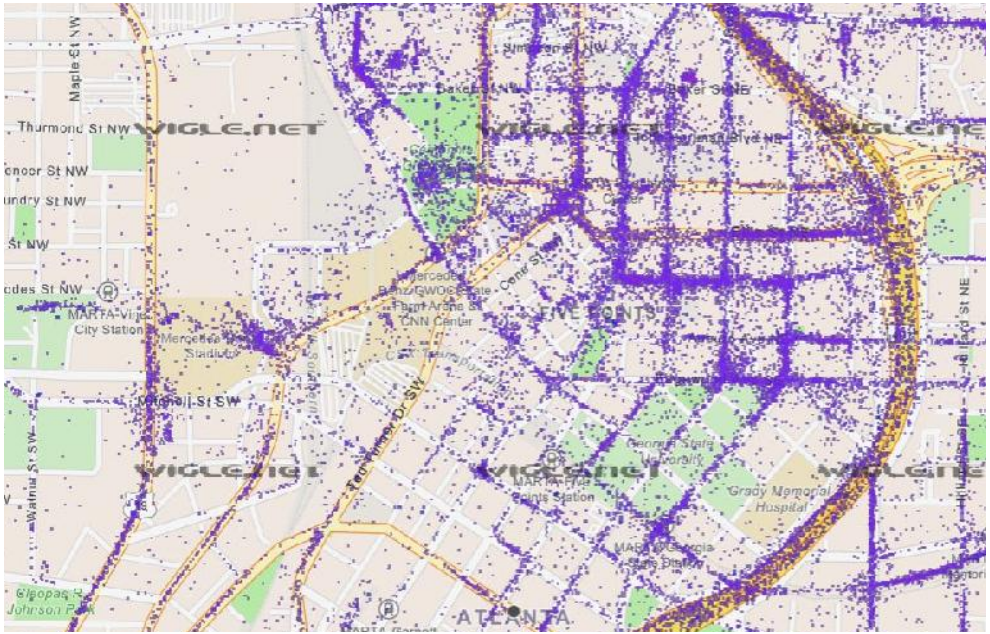
Source: https://searchengineland.com/cell-phone-triangulation-accuracy-is-all-over-the-map-14790



Source: https://www.cellmapper.net

Location sensors

- 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

Location sensors

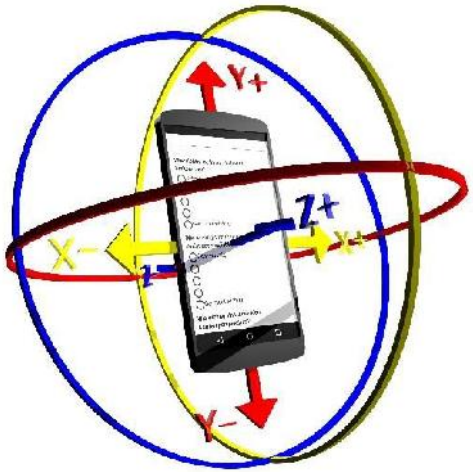
- GPS
- Cellular network
- Wi-Fi
- Hybrid positioning systems
  - Uses combination of systems to make location more accurate (assisted GPS - AGPS)
  - E.g., fall-back on X if Y is not available

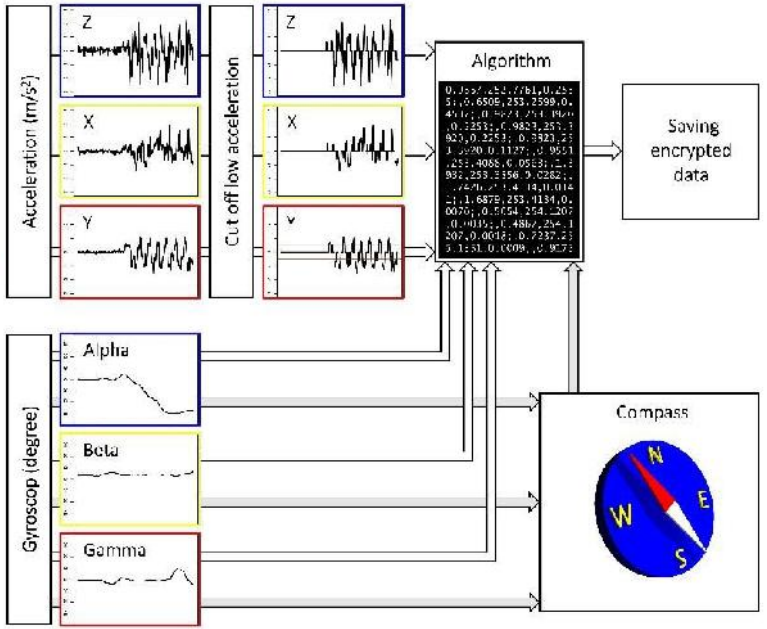
Physical activity sensors

- Accelerometer
- Gyroscope

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

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





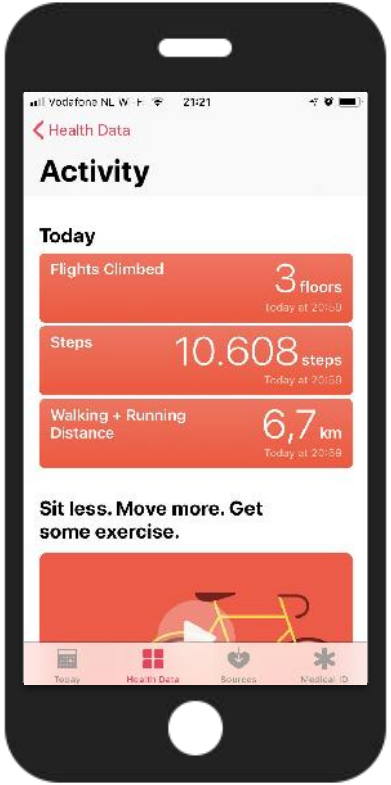
Source: Schlosser et al. (2019)

Physical activity sensors

- Accelerometer
- Gyroscope

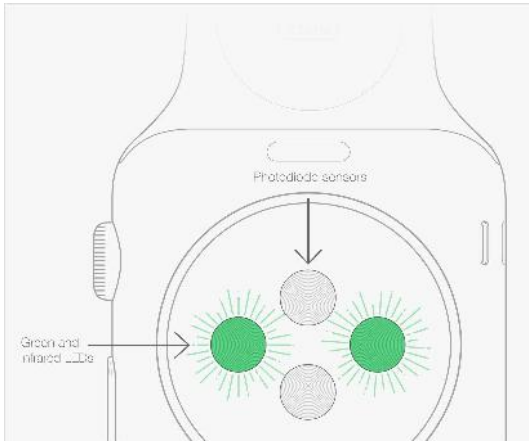
  and

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

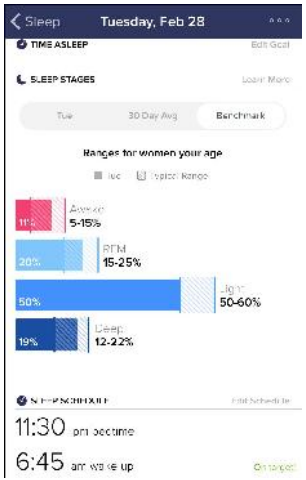


Heart-rate sensor

- 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

Microphone & light sensor

- 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 of speech 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

Smartphone-mediated behavior

- Activities inherent to functions of smartphone can be captured via use logs on device's OS
  - e.g., phone calls, text messages, app use, Internet browsing behavior, setting changes
  - Logs usually include information about type of activity, time, and duration - NO information about content

- What actually can be recorded depends on OS and user settings
  - iOS much more restrictive than Android

Camera

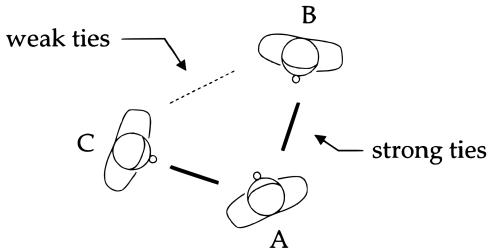
- Photos
  - Food, receipts, physical surroundings, etc.
- Video
- Barcodes
- Linear distance (iPhone Measure app)

Source: Jäckle et al. (2018)

Bluetooth, RFID, & NFC

- Bluetooth
  - Short-range communication between devices up to 30 m
  - Enabled other devices can connect to smartphones or other hubs to transmit data (e.g., weight, blood pressure, etc.)
- Beacons = small Bluetooth transmitters
  - Need to be dispatched by researcher & Bluetooth needs to be activated on receiving device
  - Great for indoor tracking
- Radio-frequency identification (RFID)
  - Electromagnetic fields to automatically identify and track tags attached to objects ~1m (3 feet)
  - e.g., assembly lines, merchandise in warehouses, livestock
- Near-field communication (NFC)
  - Communication between devices by bringing them within 4cm (1.6 in) of each other
  - e.g., contactless payment, data transfer, key cards
- All of them can be used to measure “social ties”

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



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

What kind of research questions can be answered?

Methodological … self-report & sensor data

- One source *verifies* other
  - e.g., check and edit locations and modes of transportation registered for trip by app (Scherpenzeel 2017)
  - e.g., confirming self-reported social media use through tracked app usage
- One source provides *context* for other
  - e.g., ask about reason for automatically detected trip (Green et al. 2016)
  - e.g., EMAs about happiness contextualized through information about GPS location (MacKerron & Mourato 2013)
- *Correlating* self-report with sensor data
  - e.g., self-reported stress correlates with passively measured sleep (Wang et al. 2014)
- Sensed behavior or state *triggers* survey questions
  - e.g., geofencing (Kreuter et al. 2018)
  - e.g., call to/from new number triggers questions about call (Sugie 2018)

Substantive … a selection of research questions

- How does self-reported versus objective physical activity vary across countries? (Kapteyn et al. 2018)
- Do social connections influence health and well being? (Fingerman et al. 2019)
- How do environmental factors affect happiness? (MacKerron & Mourato 2013)
- How do people interact in large social networks? (Stopczynski et al. 2014)
- How much do households spend on goods and services? (Jäckle et al. 2019; Wenz et al. 2019)
- What food and drinks do Americans acquire? (Yan et al. 2019)
- Does smartphone use correlate with personality? (Stachl et al. 2017)
- Does mental health of students change over the course of a term? (Wang et al. 2014)
- How do people find work after prison? (Sugie 2018)

Case study: IAB-SMART - Unemployment research

1930ies

Source: Archives for the History of Sociology in Austria (Graz), »Marienthal« Virtual Archives

1980ies

Source: https://tilda.tcd.ie/about/project-description/data-collection/

2010s

Source: IAB

IAB-SMART App (Kreuter et al. 2018)

- Android app, that…
  - …launches surveys
  - …passively collects smartphone data
- Over six months of data collection
  - Incentive experiment: 60 – 100€ (Haas et al. in press)
- Collected data can be combined with…
  - …German panel data (PASS)
  - …administrative records
- New data sources for unemployment research
  - Marienthal 2.0 (based on Jahoda, Lazarsfeld, & Zeisel 1933)
  - Social network analysis
- App developed by P3 insight

PASS (Trappmann et al. 2019)

- Panel study ‘Labour Market and Social Security‘
  - Household panel survey by IAB
  - Major data source for research into unemployment & poverty
- Dual frame
  - Welfare recipients from national registers: Refreshed yearly by new entries
  - General population sample from municipal registers
- ~15.000 persons in ~10.000 households each year since 2007
- Sequential mixed-mode design: CAPI -> CATI
- Main topics: labor market participation, job search, benefit receipt, active labor market programs, social inclusion, health, income, deprivation

Sample

- Sampling from PASS panel participants (aged 18-64)
- Wave 11, 2017:
  - Do you own a smartphone?: 84% YES
  - Which operating system do you use?: 70% Android
- Limited to smartphone owners with Android operating system
  - Passive access to sensor data only possible with Android
- Benefits of using PASS
  - Evaluation and separation of coverage-, nonresponse-, and measurement error
  - Higher willingness for cooperation

Communication with participants

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

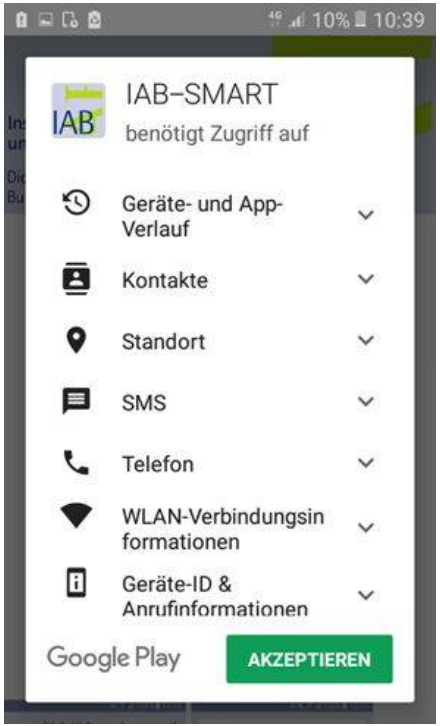
Types of data collected

- Network quality and location information (every half hour)
- Interaction history
- Characteristics of the social network
- Activity data (every two minutes)
- Smartphone usage

- Self-reports (16 survey modules)
  - Time triggered
  - Location triggered (“geofencing”)

Consent process

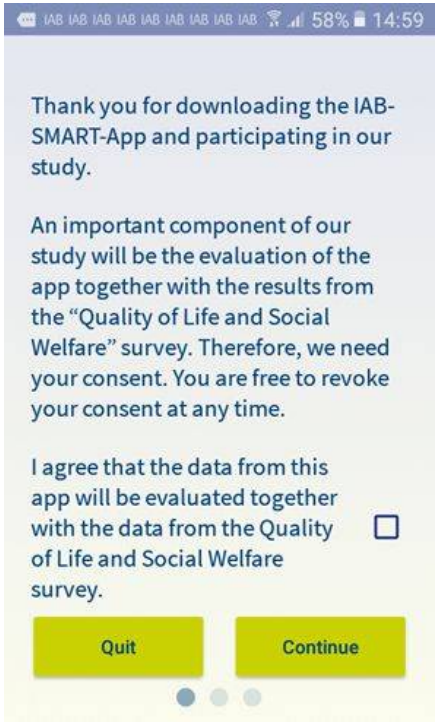
Google Play Store



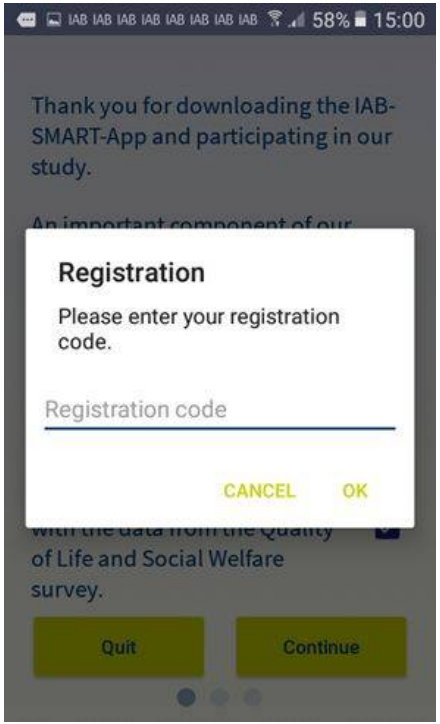
Google Permissions

App installed

Consent process



Linkage to PASS



Registration

Data processing

Consent process

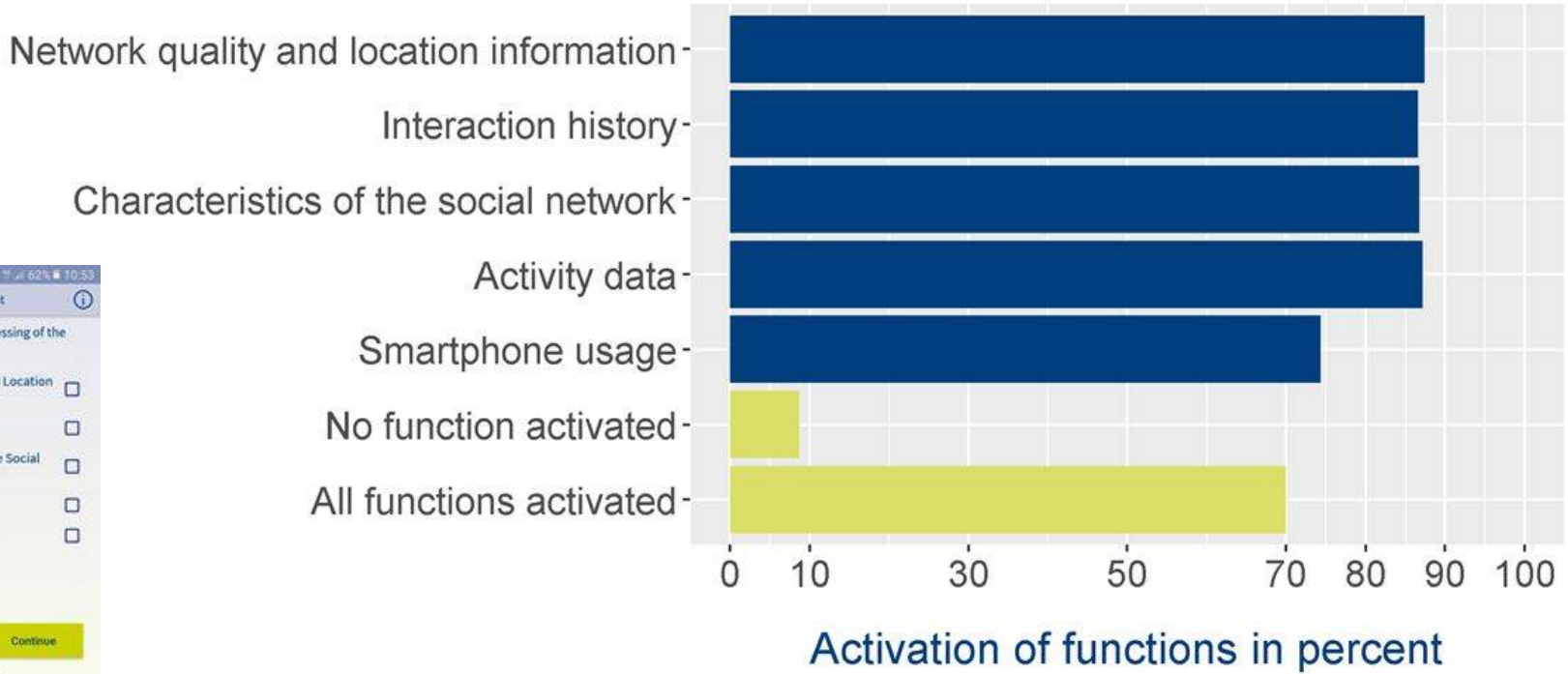
Individual consent screen

Function explanation

Full consent

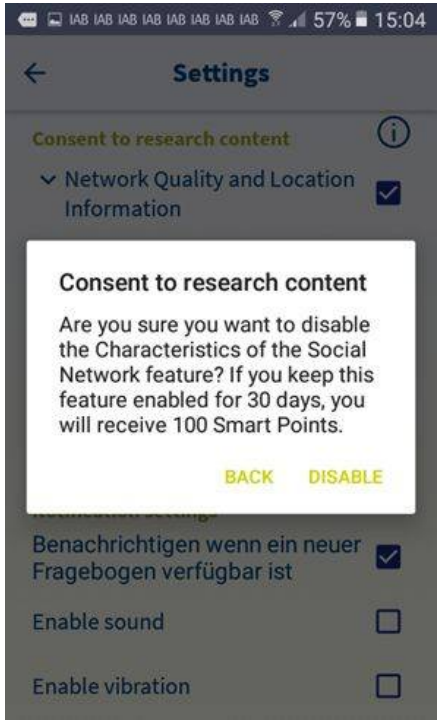
App home screen

Consenting to different functions

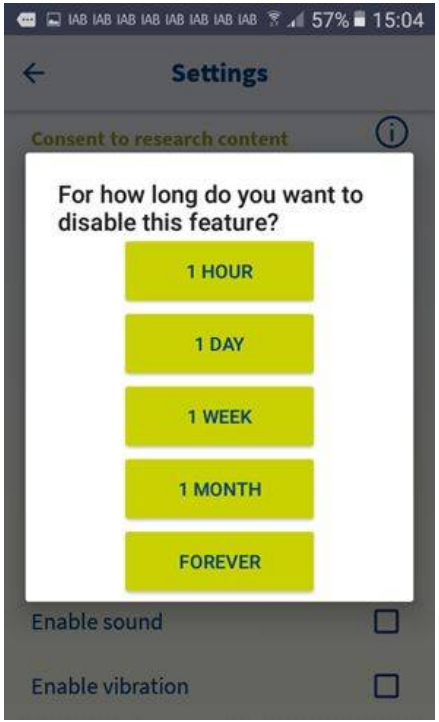


Withdrawing consent

App settings



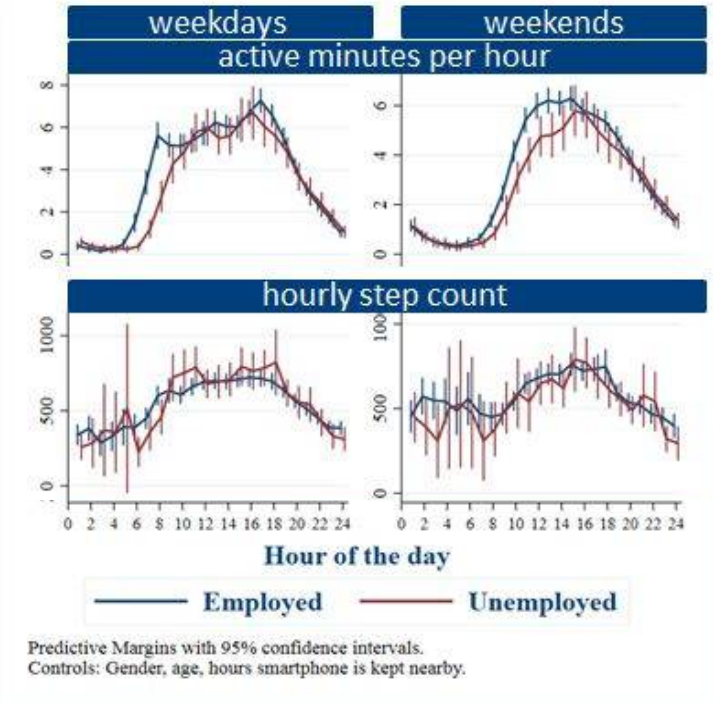
Withdrawing consent



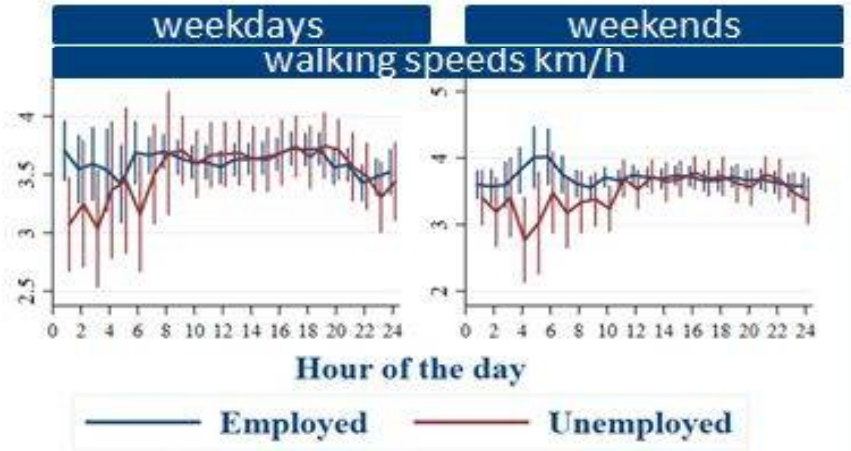
Withdrawing consent

App settings

What are the effects of unemployment? (Bähr et al. 2018)



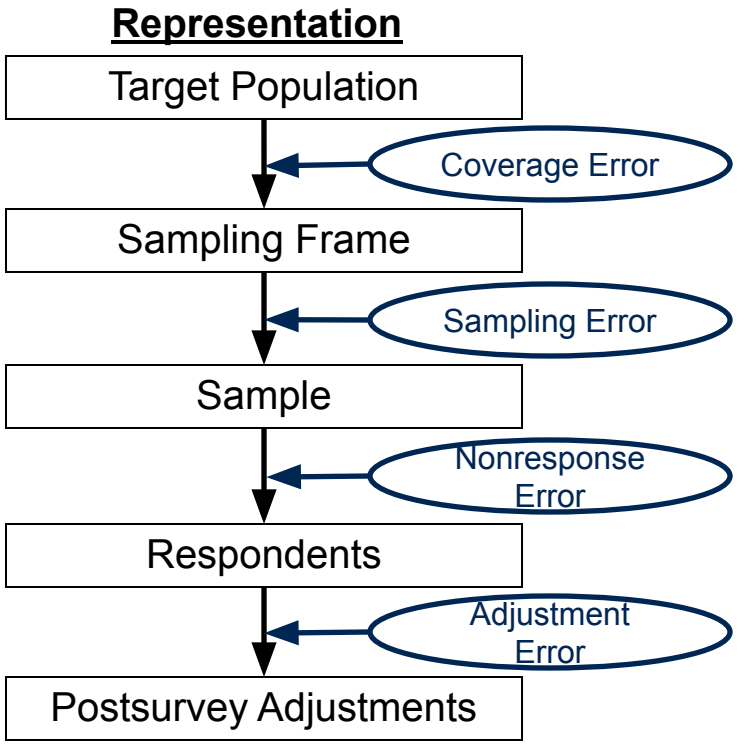
Predictive Margins with 95% confidence intervals.
Controls: Gender, age, hours smartphone is kept nearby.



Predictive Margins with 95% confidence intervals.
Controls: Gender, age, hours smartphone is kept nearby.

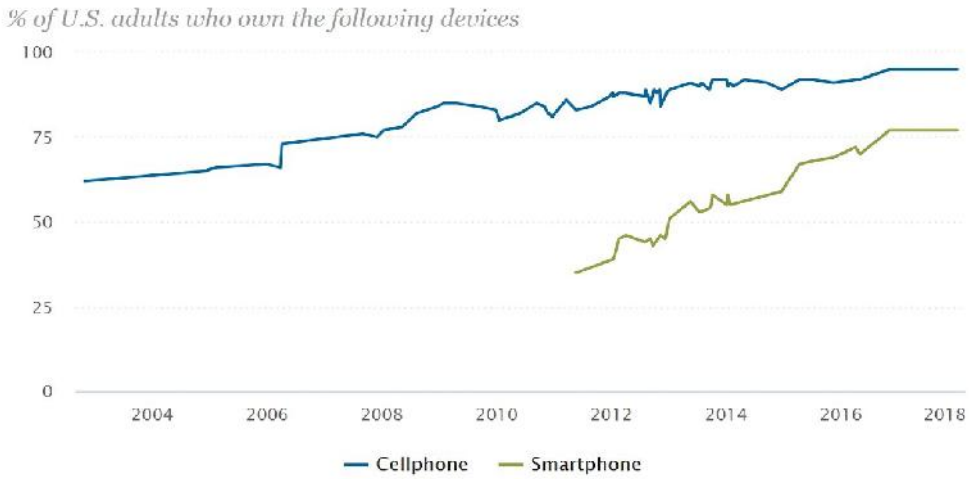
Study design considerations from the TSE perspective

Examples of representation error in app, sensor, and 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.
- **Nonresponse 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.

Coverage smartphones

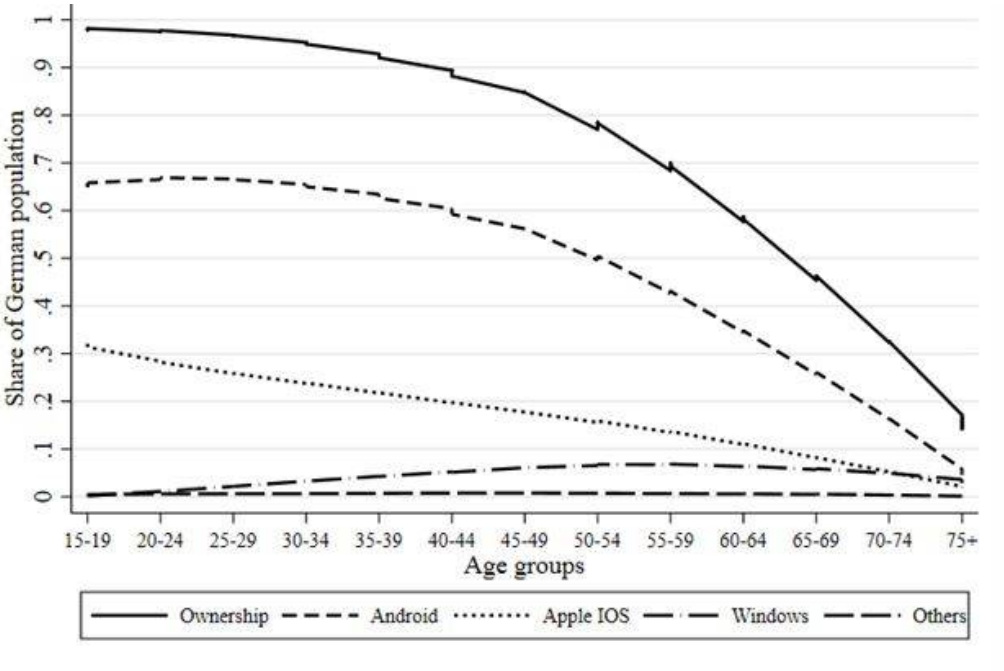


Source:https://www.pewinternet.org/fact-sheet/mobile/

- In Europe, mobile Internet access varies between 31% (Italy) and 84% (Netherlands, Sweden) in 2017 (Eurostat 2018)
- Comparable numbers for Asian-Pacific area (Silver 2019)
- Levels in Africa substantially lower with much variability across countries (Silver 2019)

Smartphone coverage bias in Germany

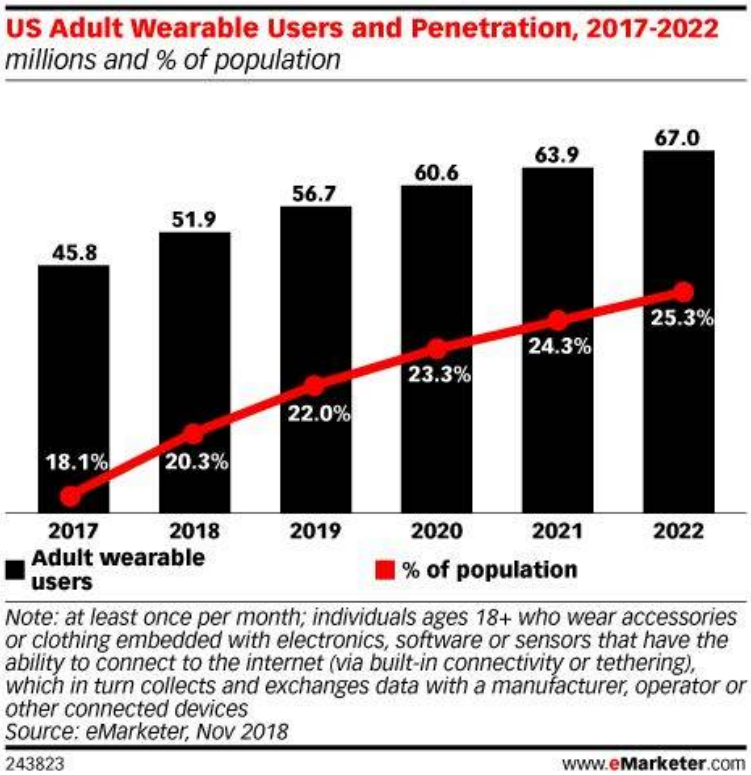
(Keusch et al. under review)



- Smartphone ownership also correlates with…
  - Educational attainment
  - Nationality
  - Region
  - Community size
- Absolute bias in substantive PASS measures for smartphone ownership relatively small (< +/-6 p.p.)
  - Can be reduced by weighting for sociodemographics (< +/-3 p.p.)
- Bias varies by OS

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

Wearables coverage in the U.S.



- In 2015, 24% of those aged 25 to 34 had wearable device
  - Projection for 2019: 38%
- 6.5% of those aged 55 to 64 had one
  - Projection for 2019: 13%

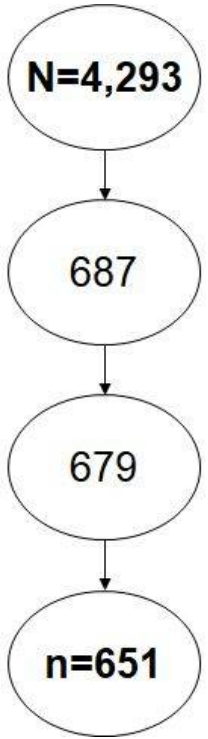
Source: https://www.emarketer.com/content/wearables-2019

Nonresponse: Willingness to participate & actual consent

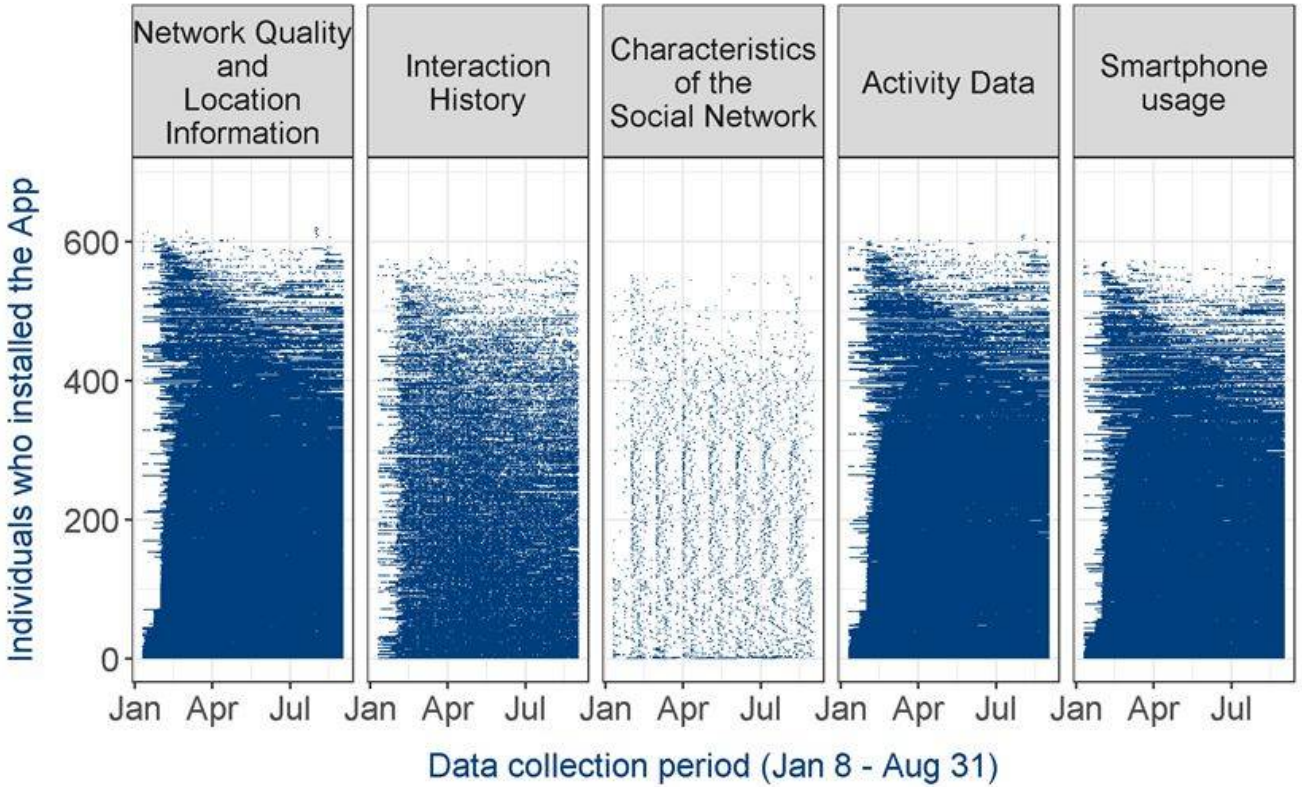
- Small-scale studies relying on enthusiasts: Willingness = 100%
  - e.g., Wang et al. (2014), Adams et al. (2014)
- Non-probability online access panels: Willingness between 5% and 56%
  - Varies across countries and by sensor type: 25%-52% for taking pictures, 19-37% for sharing GPS location (Revilla et al. 2016)
- Probability-based panels:
  - LISS Panel: Mobility (GPS, accelerometer) 37% willing, 81% participated; Physical activity (wearables) 57% willing, 90% participated (Scherpenzeel 2017)
  - UK Understanding Society Innovation Panel: Download budget app 17% (Jäckle et al. 2019)
  - German PASS: 16% installed IAB-SMART app (Kreuter et al. 2018)
- Cross-sectional general population studies:
  - CBS Travel App Download & Registration: 35% (McCool et al. 2019)
  - WTP varies from 12% for photo of house to 67% for GPS (Struminskaya et al. 2018)

Who counts as a participant? - IAB SMART

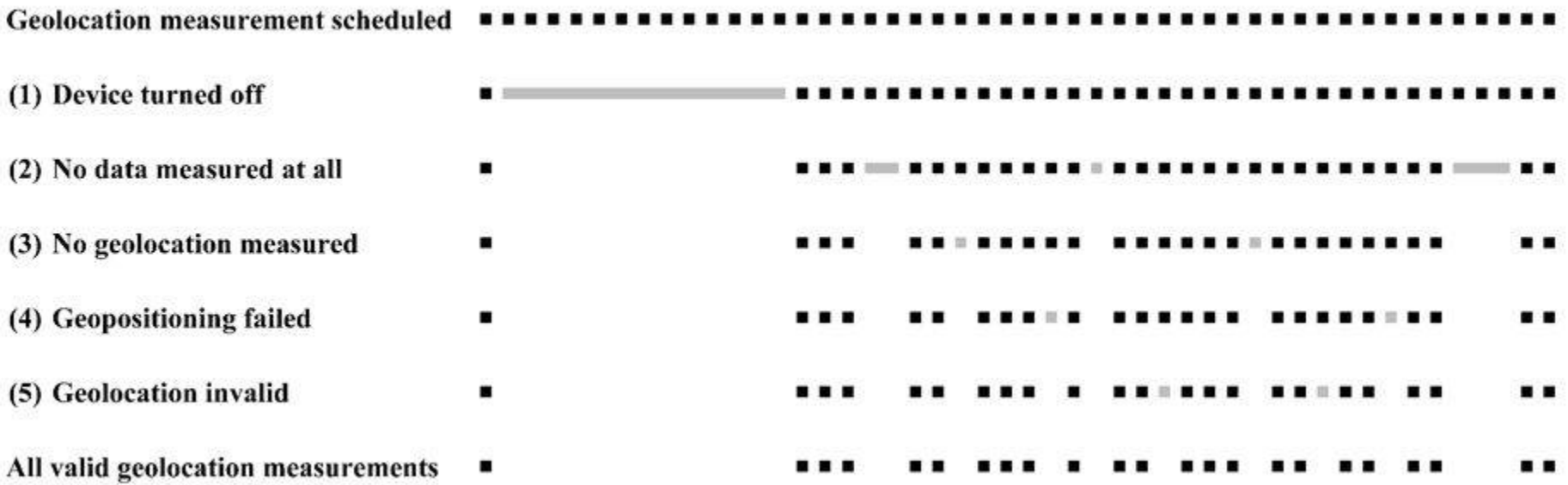
- Invited W11 PASS participants with Android smartphone
- App installations
  - Valid registration code entered in app
- Any data submitted
  - Any passive measure or answered at least one survey question
- Data from correct person
  - Age and gender in app align with PASS W11 data



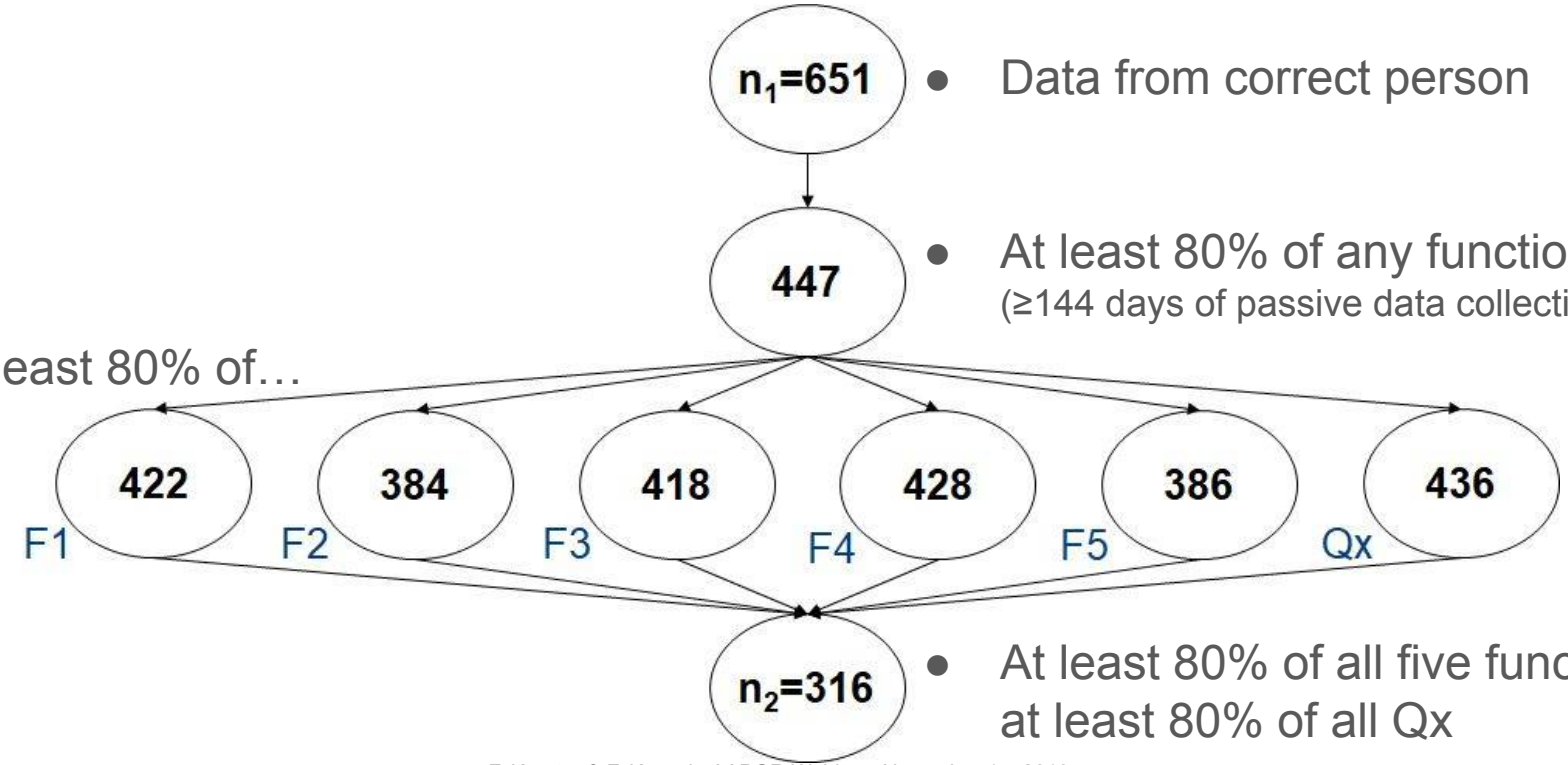
Missing data over time



Missing data over time



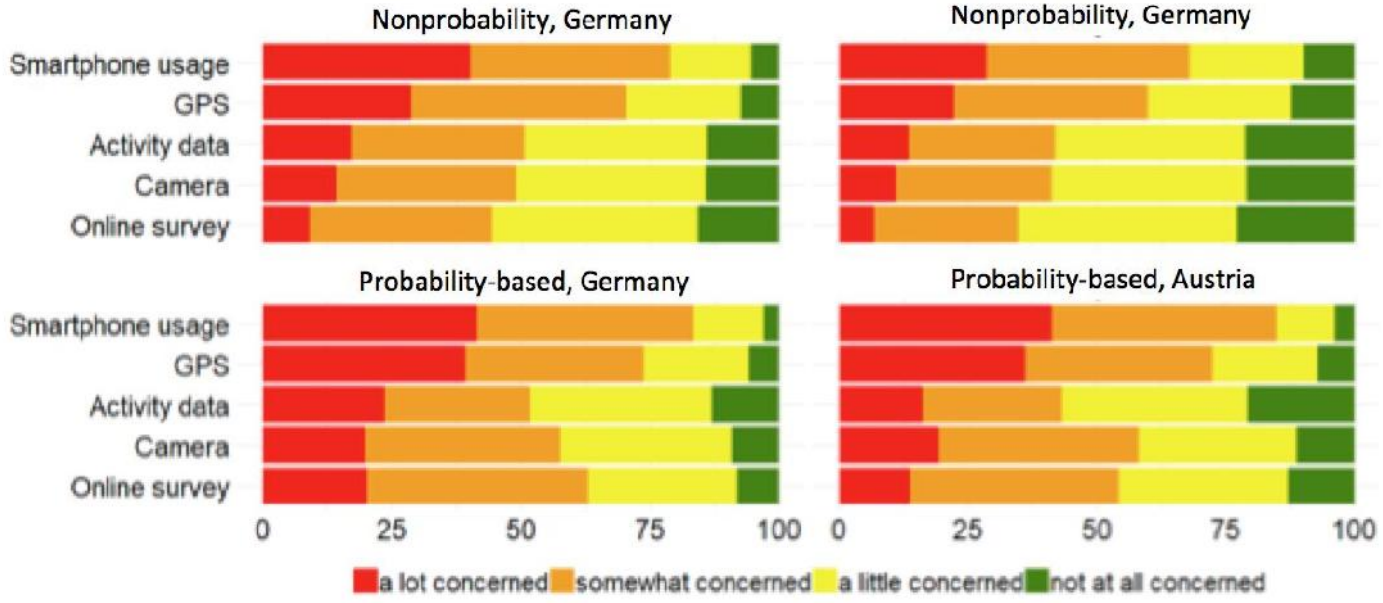
What if we use “80%-rule” from AAPOR Standard Definitions?



- Data from correct person
- At least 80% of any function
  (≥144 days of passive data collection)
- At least 80% of…
- At least 80% of all five functions + at least 80% of all Qx

Mechanisms of (non-)participation: Concern

- Privacy/security concerns: higher privacy concerns correlate with lower WTP
  (Keusch et al. in press; Revilla et al. 2018; Struminskaya et al. 2019; Wenz et al. 2019)



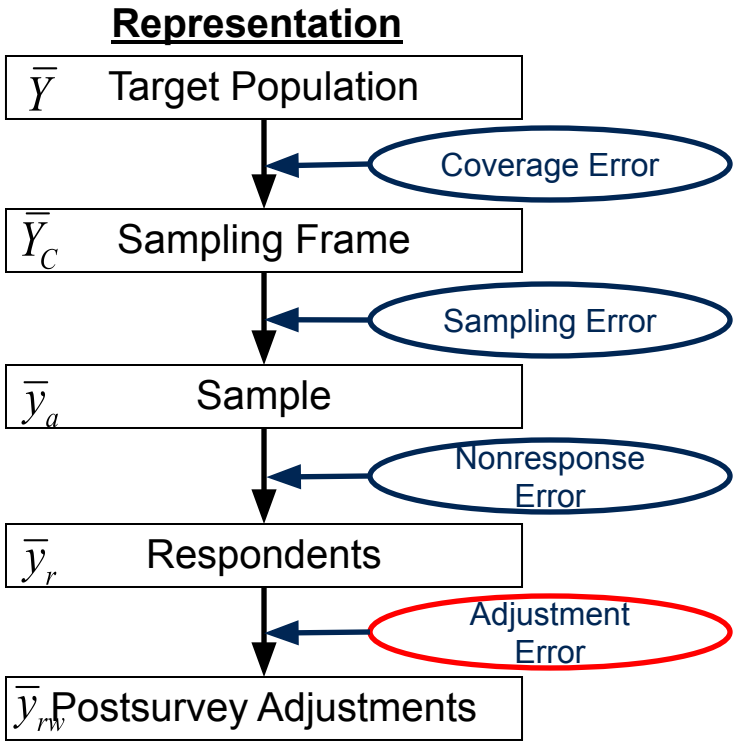
Source: Keusch et al. (in press)

Privacy & consent

- 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

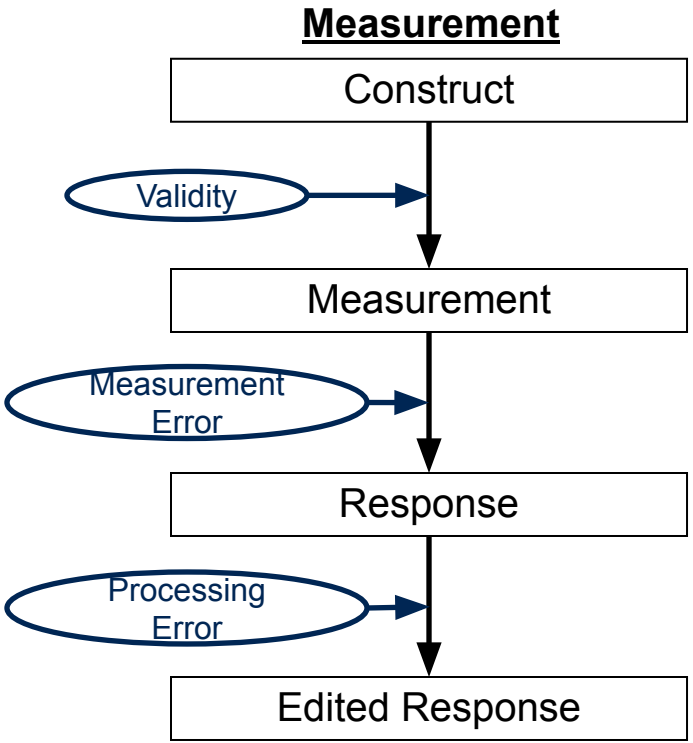
- Collecting IRB/GDPR-conforming consent
- Processing raw data on device
- Collecting data at lowest frequency necessary to answer research question

How to weight?



- **Adjustment error:** Within a study different assumptions can (need to) be made with respect to the population of inference.
  - *Smartphone mediated behavior* might only reflect the smartphone population or more specifically the given OS population
  - *Non-mediated behavior* might be more generalizable, little research has been done here

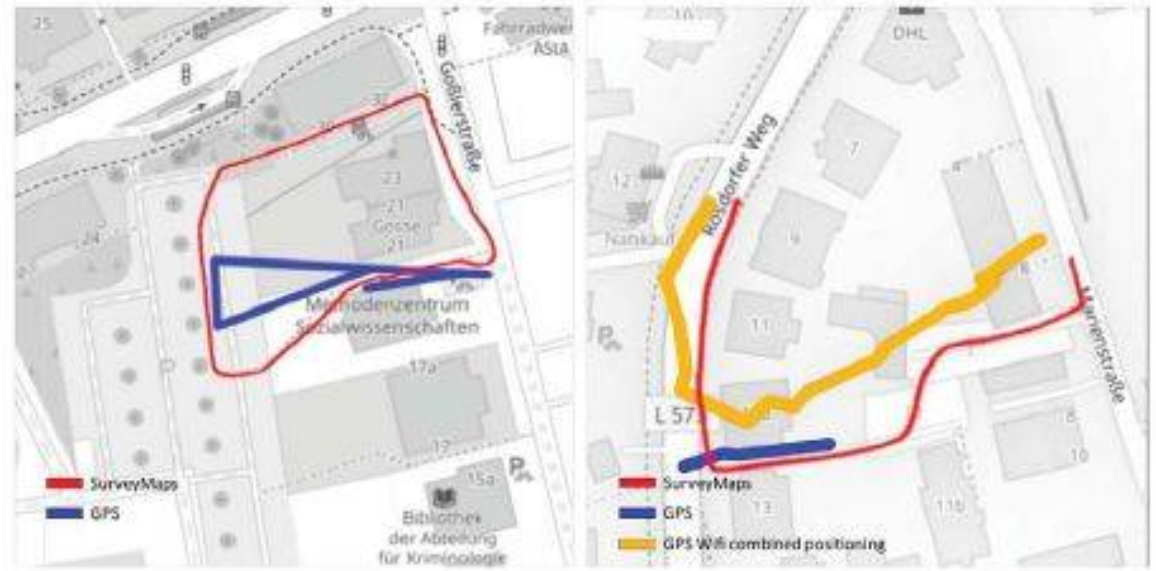
Measurement error in app, sensor, and wearables data collection



- **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).

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
- Device handling
- Erroneous data
- Providing feedback & measurement reactivity



Source: Schlosser et al. (2019)

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
  - Do people use device as anticipated by research?
- Erroneous data
- Providing feedback & measurement reactivity

Source: Sztyler et al. (2017)

Errors during data collection

- Sensor-based errors
- Device handling
- Erroneous data
  - e.g., fake GPS apps, VPN
- Providing feedback & measurement reactivity

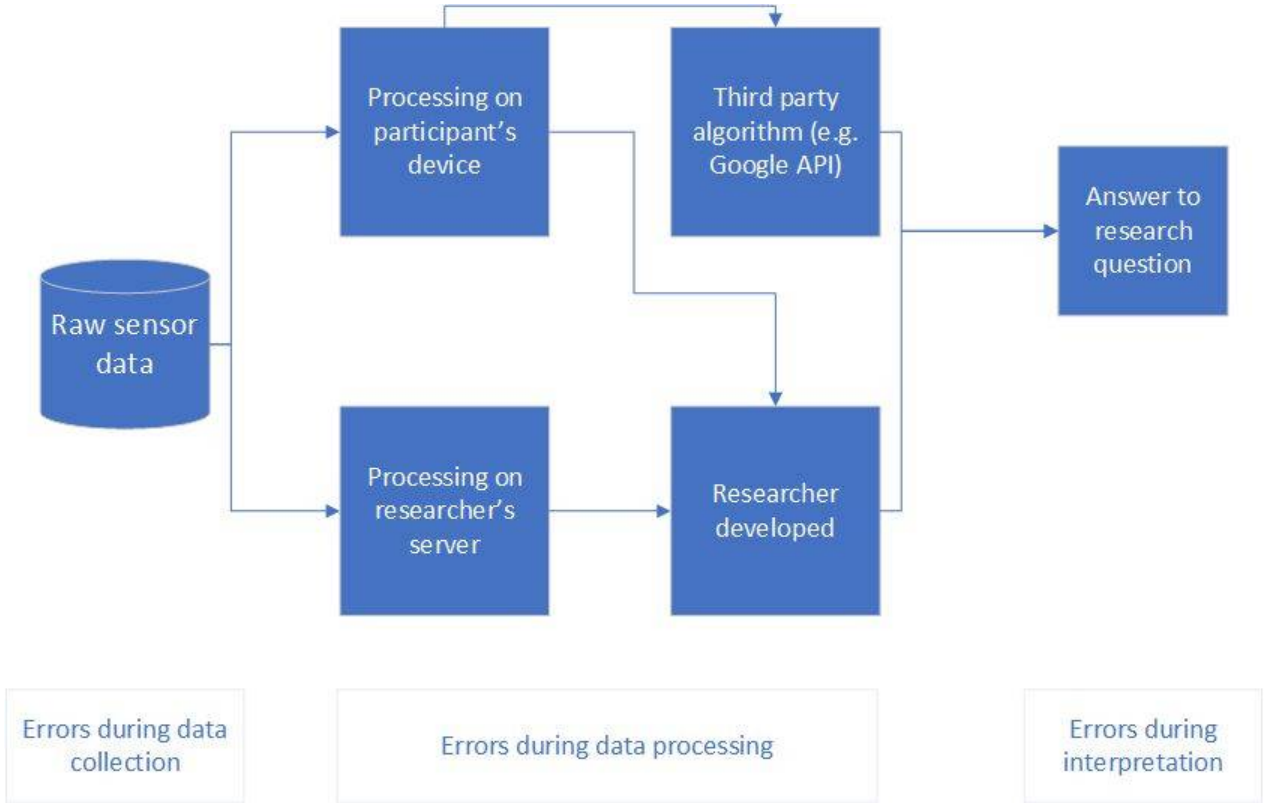
Errors during data collection

- Sensor-based errors
- Device handling
- Erroneous data
- Providing feedback & measurement reactivity
  - Has not received much attention yet

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

Processing data from sensors, apps, and wearables

From raw data to insights



F. Kreuter & F. Keusch; AAPOR Webinar; November 14, 2019

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: 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
```

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” depend 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

Model building pipeline

Data Cleaning & pre-processing

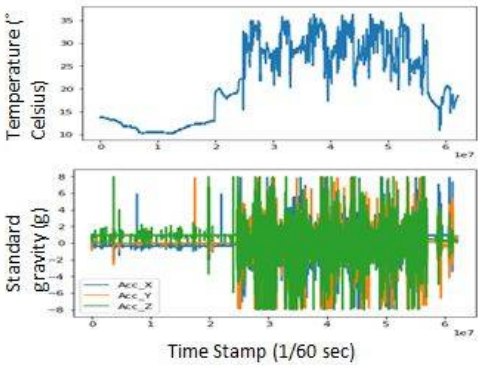
- Removal non-wear time
- Removal of high frequency (frequency higher than 15 Hz)
- Data with wear time less than 7 days discarded

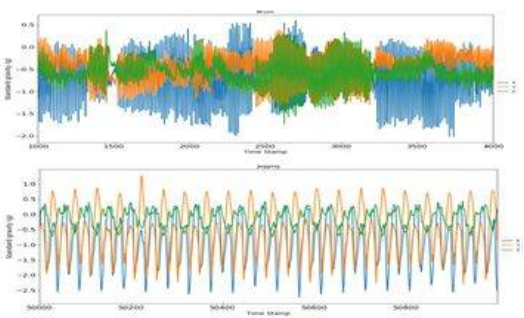
Feature Engineering

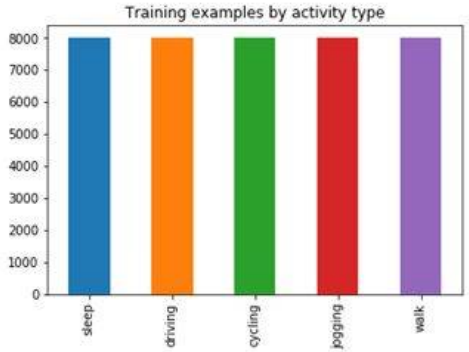
- **Time domain:** X, Y, Z, temperature, mean, median, standard deviation, RMS, percentile distribution
- **Frequency domain:** FT, dominant frequency selection, power of signal

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)



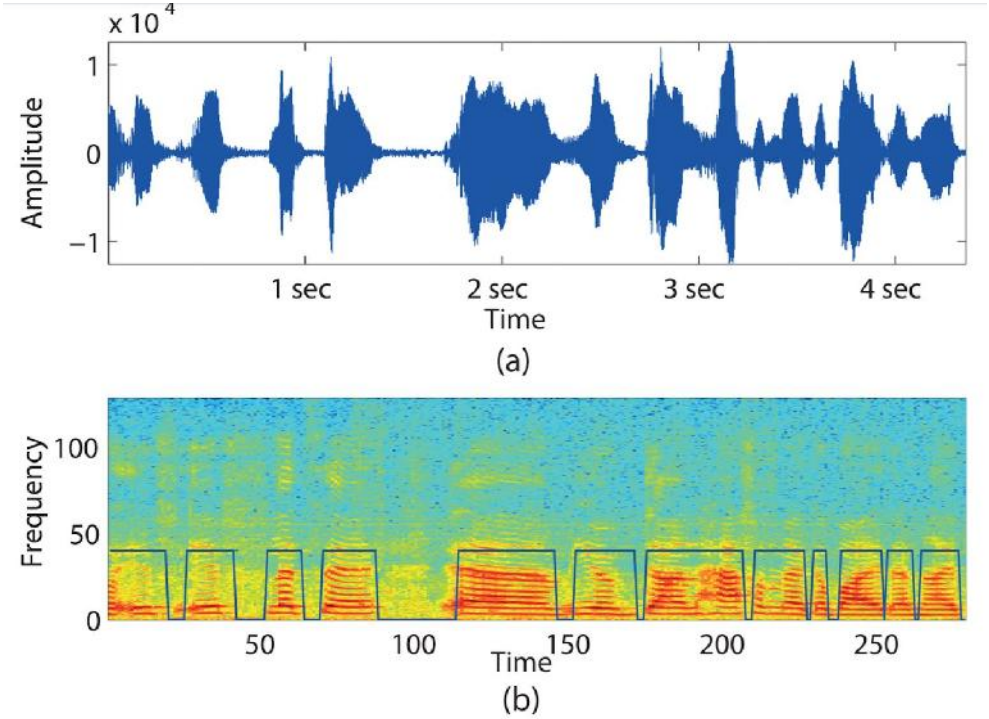




Source: Mulder et al. (2019)

Example: 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



Source: Rabbi et al. (2011)

What we talked about

- Why add apps, sensors, and wearables to surveys
- What can we measure
- What kind of research questions can be answered
  - Case Study: IAB-SMART
- Study design considerations from the TSE perspective
- Processing data from sensors, apps, and wearables

Questions

Contact

If you have questions, need more information, or want to collaborate:

fkreuter@umd.edu

f.keusch@uni-mannheim.de

References

Adams, Phil, Mashfiqui Rabbi, Tauhidur Rahman, Mark Matthews, Amy Voida, Geri Gay, Tanzeem Choudhury, and Stephen Voida. 2014. “Towards personal stress informatics: comparing minimally invasive techniques for measuring daily stress in the wild.” *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, 72-79.

Bähr, Sebastian, Georg-Christoph Haas, Florian Keusch, Frauke Kreuter, and Mark Trappmann. 2018. Marienthal 2.0: Analyzing subtle effects of unemployment using smartphones. Poster presented at the Workshop Analytische Soziologie: Theorie und empirische Anwendungen. November 12-15, Venice, Italy.

Doherty, Aiden, Dan Jackson, Nils Hammerla, Thomas Plötz, Patrick Olivier, Malcolm H. Granat, Tom White et al. 2017. “Large scale population assessment of physical activity using wrist worn accelerometers: the UK biobank study.” *PloS One* 12(2):e0169649.

Eurostat. 2018. Digital Economy and Society: ICT Usage in Households and by Individuals. http://ec.europa.eu/eurostat/web/digital-economy-and-society/data/database.

Fingerman, Karen L., Meng Huo, Susan T. Charles, and Debra J. Umberson. "Variety is the spice of late life: Social integration and daily activity." *The Journals of Gerontology: Series B* https://doi.org/10.1093/geronb/gbz007.

Greene, Elizabeth, Leah Flake, Kevin Hathaway, and Michael Geilich. 2016. “A seven-day smartphone-based GPS household travel survey in Indiana.” Paper presented at the Transportation Research Board 95th Annual Meeting. January 10-14, Washington, DC.

Haas, Georg-Christoph, Frauke Kreuter, Florian Keusch, Mark Trappmann, and Sebastian Bähr. in press. “Effects of incentives in smartphone data collection.” In A. Kirchner, et al. (Eds.) *Big Data Meets Survey Science*. Hoboken, NJ: Wiley.

Harari, Gabriella M., Sandrine R. Müller, Min S.H. Aung, and Peter J. Rentfrow. 2017. “Smartphone sensing methods for studying behavior in everyday life.” *Current Opinion in Behavioral Sciences*, 18:83-90.

Jahoda, Marie, Paul F. Lazarsfeld, and Hans Zeisel. 1933. *Marienthal: The sociography of an unemployed community*.
Jäckle, Annette, Burton, Jonathan, Couper, Mick P., Lessof, Carli. 2019. “Participation is a Mobile App Survey to Collect Expenditure Data as Part of a Large-Scale Probability Household Panel: Coverage and participation rates and biases.” *Survey Research Methods* 13:23-44.
Kapteyn, Arie, Banks, James, Hamer, Mark, Smith, James P., Steptoe, Andrew, Van Soest, Arthur, Koster, Annemarie, and Saw Htay Wah. 2018. What They Say and What They Do: Comparing physical Activity Across the USA, England, and the Netherlands. *Journal of Epidemiological Community Health* 72:471-476.
Keusch, Florian, Sebastian Bähr, Georg-Christoph Haas, Frauke Kreuter, and Mark Trappmann. under review. “Coverage error in data collection combining mobile surveys with passive measurement using apps: Data from a German national survey.”
Keusch, Florian, Bella Struminskaya, Frauke Kreuter, and Martin Weichbold. in press. “Combining active and passive mobile data collection: A survey of concerns.” In A. Kirchner, et al. (Eds.) *Big Data Meets Survey Science*. Hoboken, NJ: Wiley.
Kreuter, Frauke, Haas, Georg-Christof, Keusch, Florian, Bähr, Sebastian, and Trappman, Mark 2018. “Collecting Survey and Smartphone Sensor Data With an App: Opportunities and Challenges Around Privacy and Informed Consent”. *Social Science Computer Review* http://doi.org/10.1177/0894439318816389.
MacKerron, George, and Susana Mourato. 2013. “Happiness is greater in natural environments.” *Global Environmental Change* 23:992-1000.
McCool, Danielle, OleMussman, Barry Schouten, Victor Verstappen, and Peter Lugtig. 2019. Statistics Netherlands Travel App. Paper presented at the 1st MASS Workshop, March 4-5, Mannheim, Germany.
Mulder, Joris,Natalia Kieruj, Seyit Höcük, and Pradeep Kumar. 2019. What really makes you move? Identifying relationships between physical activity and health through applying machine learning techniques on high frequency accelerometer and survey data. Paper presented at the 1st MASS Workshop, March 4-5, Mannheim, Germany.

Rabbi, Mashfiqui, Shahid Ali, Tanzeem Choudhury, and Ethan Berke. 2011. “Passive and in-situ assessment of mental and physical well-being using mobile sensors.” In *Proceedings of the 13th International Conference on Ubiquitous Computing*, 385-394.

Revilla, Melanie, Mick P. Couper, and Carlos Ochoa. 2018. “Willingness of Online Panelists to Perform Additional Tasks.” *Methods, data, analyses.* http://doi.org/10.12758/mda.2018.01.

Revilla, Melanie, Daniele Toninelli, Carlos Ochoa, and German Loewe. 2016. “Do Online Access Panels Need to Adapt Surveys for Mobile Devices?” *Internet Research* 26:1209-1227.

Scherpenzeel, Annette. 2017. “Mixing online panel data collection with innovative methods.” In S. Eifler & F. Faulbaum (Eds.): *Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung*. Wiesbaden: Springer.

Schlosser, Stephan, Jan Karem Höhne, and Daniel Qureshi. 2019. SurveyMaps: A sensor-based supplement to GPS in mobile web surveys. Paper presented at the first MASS workshop. March 4-5, Mannheim, Germany.

Silver, Laura. (2019, February 5). Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center’s Global Attitudes Project. https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/.

Stachl, Clemens, Sven Hilbert, Jiew-Quay Au, Daniel Buschek, Alexander De Luca, Bernd Bischl, Heinrich Hussmann, and Markus Bühner. “Personality traits predict smartphone usage.” *European Journal of Personality* 31(6):701-722.

Stopczynski, Arkadiusz, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, Mette My Madsen, Jakob Eg Larsen, and Sune Lehmann. 2014. “Measuring large scale social networks with high resolution.” *PLOS One* 9(4):e95978.

Struminskaya, Bella, PeterLugtig, Barry Schouten, and Vera Toepoel. 2019. Willingness to collect smartphone-based measurements: Asking for consent. Paper presented at the first MASS workshop. March 4-5, Mannheim, Germany.

Sugie, Naomie F. 2018. Utilizing smartphones to study disadvantaged and hard-to reach groups. *Sociological Methods and Research* 47:458-491.

Trappmann, Mark, Sebastian Bähr, Jonas Beste, Andreas Eberl, Corinna Frodermann, Stefanie Gundert, Stefan Schwarz, Nils Teichler, Stefanie Unger, and Claudia Wenzig. 2019. “Data resource profile: Panel Study Labour Market and Social Security (PASS).” *International Journal of Epidemiology* https://doi.org/10.1093/ije/dyz041.

Wang, Rui, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. 2014. “StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones.” *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 3-14.

Wenz, Alexander, Annette Jäckle, and Mick P. Couper. 2019. “Willingness to use mobile technologies for data collection in a probability household panel.” *Survey Research Methods* 13:1-22.

Yan, Ting, Machado, Janice, Heller, Andrew, and Marcelo Simas. 2019. The feasibility of using smartphones to record food purchase and acquisition. Paper presented at the first MASS workshop. March 4-5,Mannheim, Germany.

Sources of photos on slide 14

s1.https://www.androidcentral.com/samsung-galaxy-s
s2.https://www.samsung.com/uk/smartphones/galaxy-s2/GT-I9100LKAXEU/
s3. https://www.samsung.com/uk/smartphones/galaxy-s3-i9300/GT-I9300ZKDBTU/
s4. https://www.samsung.com/us/mobile/phones/galaxy-s/samsung-galaxy-s4-cricket-black-mist-sgh-i337zkzaio/
s5. https://www.samsung.com/uk/smartphones/galaxy-s5-g900f/SM-G900FZKABTU/
s6. https://www.samsung.com/levant/smartphones/galaxy-s6-edge-g925f/
s7. https://www.verizonwireless.com/smartphones/samsung-galaxy-s7/
s8. https://www.samsung.com/us/mobile/phones/galaxy-s/galaxy-s8-active-64gb--at-t--meteor-gray-sm-g892azaaatt/
s9. https://www.samsung.com/us/smartphones/galaxy-s9/buy/v2/
s10. https://www.samsung.com/uk/smartphones/galaxy-s10/buy/