Using Smartphones for Data Collection

Alexander Wenz

University of Mannheim

MZES Social Science Data Lab
April 6, 2022

Why collect data with smartphones?

- ► What type of data can we collect?
- ► What are the benefits compared to existing methods?

What kind of research questions can we answer?

What is the quality of smartphone data collection?

- ► Representation error
- ► Measurement error

Why collect data with smartphones?

Smartphone use in everyday life

- ▶ 76% of the general population aged 15+ in Germany own a smartphone (Keusch et al., 2020)

- ▶ 83% of smartphone users take their smartphone when leaving their home (Keusch, Wenz, & Conrad, 2022)

- ▶ Required for an increasing array of everday activities
  - ▶ Communication
  - ▶ Online banking, mobile payment
  - ▶ Mobility
  - ▶ Contact tracing, vaccination certificate
  - ▶ Hub for Internet-of-Things devices

Smartphones as tools for data collection

1. Administer web surveys

2. Collect data from the operating system and built-in sensors
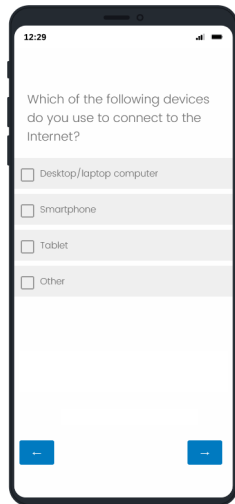
Survey-based data collection

Browser- or app-based surveys

Diary studies
- ▶ Capture events when they occur
- ▶ E.g., time use, travel, expenditure

Experience sampling
- ▶ *Ecological Momentary Assessment*
- ▶ Survey prompted at random times, regular intervals, or based on events
- ▶ E.g., subjective well-being

Survey-based data collection

Potential benefits

- ▶ *In situ* measurement of attitudes and/or behaviors
  - ▶ Reduced recall errors
  - ▶ More detailed information

- ▶ Longitudinal research design
  - ▶ Observe change over time
  - ▶ Notifications to increase compliance

Survey-based data collection

How to collect data

▶ Browser-based smartphone surveys can be programmed with common web survey software (e.g., Qualtrics, Unipark)

▶ Software for implementing survey apps and experience sampling (e.g., Samply, Murmuras)

Survey-based data collection

Design principles for smartphone questionnaires (Antoun et al., 2018)

1. Readability
   ▶ Text should be large enough to promote easy reading

2. Ease of selection
   ▶ Touch targets should be large enough to tap accurately

3. Visibility across the page
   ▶ All content should fit the screen width, no horizontal scrolling

Survey-based data collection

Design principles for smartphone questionnaires (Antoun et al., 2018)

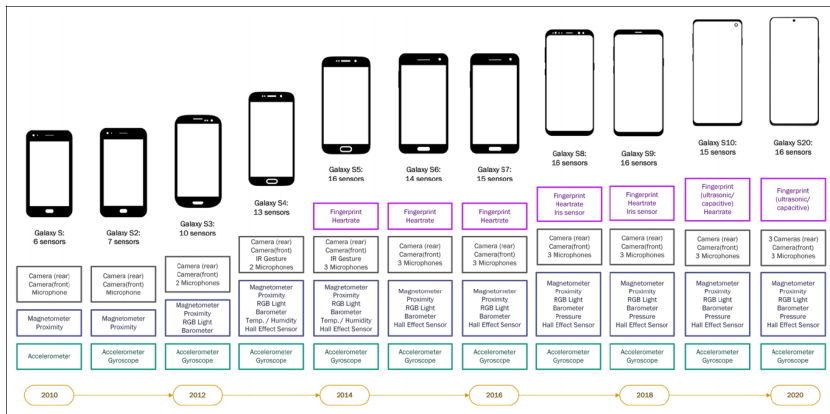4. Simplicity of design features
   - ▶ Simple for researchers to deploy and for respondents to use

5. Predictability across devices
   - ▶ Questionnaire design should be predictable across devices

Smartphones as tools for data collection

1. Administer web surveys

2. Collect data from the operating system and built-in sensors
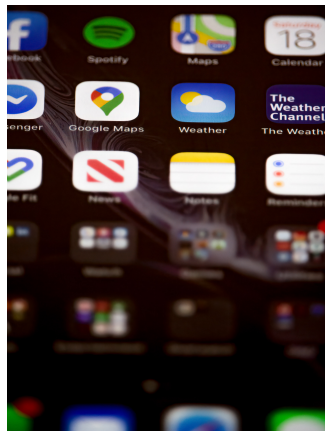
# Sensor-based data collection

Source: Struminskaya et al. (2020)

UNIVERSITY
OF MANNHEIM
School of Social Sciences

Smartphone behavior
Operating system log files

- ▶ Phone calls
- ▶ Text messages
- ▶ App use
- ▶ Internet browsing behavior

Sensor-based data collection

Smartphone behavior

- ▶ Information about type of activity, time, and duration – usually no information about content
- ▶ What actually can be recorded depends on the operating system and user settings
- ▶ iOS much more restrictive than Android

Sensor-based data collection
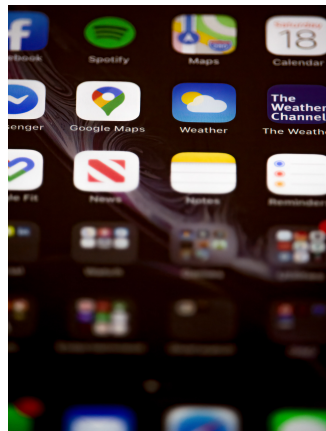
Location and Movement
GPS

- ▶ Captures coordinates in latitude and longitude based on distance to 4+ satellites

- ▶ e.g. University of Mannheim: (49° 28' 35.70'' N, 8° 27' 31.72'' E)

- ▶ Accuracy within 30 centimeters

- ▶ Works without cell connection

- ▶ Performs worse in urban canyons, indoors, and underground

- ▶ Very battery-draining

Sensor-based data collection

Location and Movement

Cellular network

- ▶ Based on radio signals between several cell towers
- ▶ High accuracy in urban areas

Wi-Fi

- ▶ Infers location from Wi-Fi access points
- ▶ Can overcome problems of urban canyons and indoor tracking

Sensor-based data collection

Physical activity

Accelerometer

▶ Measures acceleration on 3 axes

Gyroscope

▶ Measures if device is tilted

Magnetometer

▶ Serves as compass

Barometer

▶ Tracks changes in elevation

(Höhne & Schlosser, 2019)

Proximity

Bluetooth

▶ Short-range communication
between devices up to 30 meters

▶ Can connect to small transmitters
(Beacons) or other devices to
transmit data (e.g., smart scales)
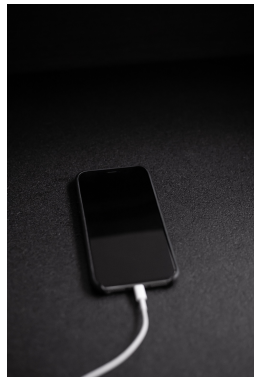
Sensor-based data collection

Sleep

Microphone

- ▶ Measures ambient noise
- ▶ Classifiers determine that participant is in noisy vs. silent environment but do not reconstruct content of recording

Light sensor

Accelerometer

Operating system log files

Sensor-based data collection

Potential benefits

- ▶ Passive measurement of behaviors
    - ▶ No self-report = No recall error
    - ▶ No self-report = Potentially less social desirability
    - ▶ More accurate data (not always!)

- ▶ More detailed data
    - ▶ High measurement frequency (e.g., accelerometer with 60 Hz)
    - ▶ Information that cannot be self-reported (e.g., stages of sleep)
    - ▶ Various data formats from one device (e.g., location, pictures)

- ▶ Less response burden
    - ▶ Fewer survey questions have to be answered
    - ▶ But what about other burden, e.g., consent, compliance?

Sensor-based data collection

How to collect data

▶ Develop your own app
  ▶ e.g., ResearchStack for Android, Apple Research Kit
  ▶ But very labor-intensive!

▶ Use data collection apps
  ▶ e.g., Murmuras, Ethica (can track smartphone behavior, location, physical activity, proximity, etc.)

▶ Use JavaScript-based tools
  ▶ e.g., SurveyMotion, SurveyImage, SurveyVoice (can track physical activity, collect images, record voice answers in web surveys)

What kind of research questions can we answer?

Methodological research questions

One source **verifies** other

▶ Check locations and modes of transportation captured by app
(Scherpenzeel, 2017)

▶ Confirm self-reported Internet use through tracked Internet use
(Scharkow, 2016)

One source **provides context** for other

▶ Ask about reason for automatically detected trip
(Greene et al., 2016)

▶ EMA about happiness contextualized through information about GPS location (MacKerron & Mourato, 2013)

Methodological research questions

Correlate self-report with sensor data

▶ Self-reported stress correlates with passively measured sleep
(Wang et al., 2014)

Sensed behavior or state **triggers** survey questions

▶ Geofence around job centers triggers survey about consulting meeting (Haas et al., 2020)

▶ Call to/from new phone number triggers question about the call (Sugie, 2018)

Substantive research questions

▶ How does unemployment affect people's everyday life?
(Kreuter et al., 2020)

▶ How do people find work after prison? (Sugie, 2018)

▶ How does the sociability of students change over the course of a term? (Harari et al., 2020)

▶ Do social connections influence health and well-being?
(Fingerman et al., 2020)

▶ What is the relationship between physical activity and happiness? (Lathia et al., 2017)

▶ How much do people spend on goods and services?
(Jäckle et al., 2019)

What is the quality of smartphone data collection?

Quality of smartphone data collection

1. Representation error
2. Measurement error

Representation error

- ▶ Most previous research based on volunteer samples
    - ▶ e.g., Wang et al. (2014) n=48 students

- ▶ Low participation rates in general population samples
    - ▶ UKHLS Spending Study 1: 13% (Jäckle et al., 2019)
    - ▶ IAB-SMART Study: 16% (Kreuter et al., 2020)

- ▶ Barriers to participation
    - ▶ Smartphone access
    - ▶ Smartphone skills
    - ▶ Willingness to use smartphone for data collection
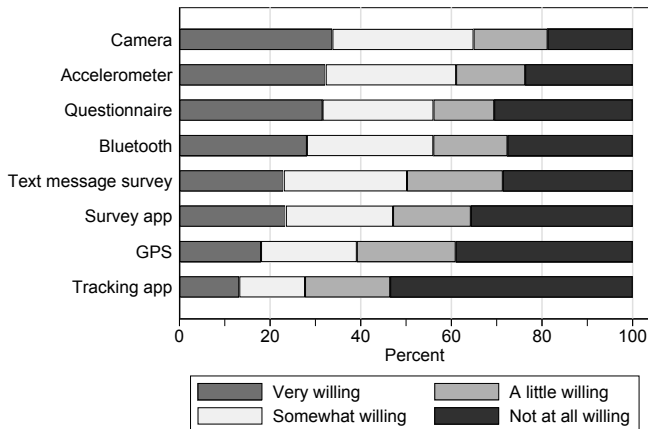
Smartphone access

- ▶ 76% of people aged 15+ in Germany own a smartphone
- ▶ Smartphone ownership correlates with
    - ▶ Age
    - ▶ Educational attainment
    - ▶ Household income
- ▶ Bias can be reduced by weighting

(Source: Keusch et al., 2020)
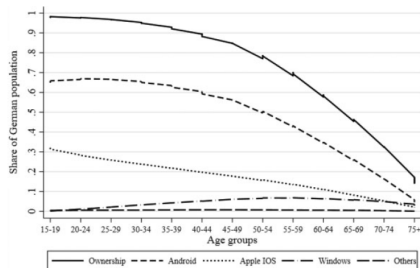
Willingness to use smartphone

▶ Willingness varies by type of data collected (Wenz et al., 2019)

Note. Questionnaire, survey app, camera, accelerometer: N = 1,376.
Text message survey, GPS, bluetooth: N = 1,375. Tracking app: N = 1,374

Willingness to use smartphone

Respondent characteristics

- ▶ Privacy and security concerns (−)
- ▶ Trust in data collection agency (+)
- ▶ Smartphone experience (+)

Study characteristics

- ▶ Control over data collection (+)
- ▶ Study invitation within face-to-face interview (+)

Privacy and consent

- ▶ Participants might have concerns about potential risks related to sensor data
  - ▶ Data streams could be intercepted by unauthorized party
  - ▶ Connecting multiple data streams could re-identify previously anonymous user
  - ▶ Information could be used to impact credit, employment, or insurability

- ▶ Collecting GDPR-conforming consent
  - ▶ Inform participants fully about what will be collected, people can revoke permission to collect data

- ▶ Collecting data at lowest frequency necessary to answer research question

- ▶ Processing raw data on device

Quality of smartphone data collection

1. Representation error
2. Measurement error

Measurement error

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

- ▶ Device handling
    - ▶ Measurement might differ depending on how the device is worn, e.g. differences in how men and women carry around smartphones
    - ▶ Do participants use the device as anticipated by the researcher?

- ▶ Erroneous data
    - ▶ e.g., VPN

- ▶ Providing feedback and measurement reactivity

Resources

Mobile web surveys

▶ Couper, M. P., Antoun, C., & Mavletova, A. (2017). Mobile web surveys. In P. P. Biemer et al. (Eds.), *Total Survey Error in Practice* (pp. 133-154). Wiley.

Experience sampling

▶ Myin-Germeys, I., & Kuppens, P. (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.

Sensor-based data collection

▶ Harari, G. M., Müller, S. R., Aung, M. S., & Rentfrow, P. J. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences, 18*, 83-90.

▶ Keusch, F. & Conrad, F. G. (2021). Using smartphones to capture and combine self-reports and passively measured behavior in social research. *Journal of Survey Statistics and Methodology*, 1-23.

Questions?

References

► Antoun, C., Katz, J., Argueta, J., & Wang, L. (2018). Design heuristics for effective smartphone questionnaires. *Social Science Computer Review, 36*(5), 557-574.

► Fingerman, K. L., Huo, M., Charles, S. T., & Umberson, D. J. (2020). Variety is the spice of late life: Social integration and daily activity. *The Journals of Gerontology: Series B, 75*(2), 377-388.

► Greene, E., Flake, L., Hathaway, K., & Geilich, M. (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, G.-C., Trappmann, M., Keusch, F., Bähr, S., & Kreuter, F. (2020). Using geofences to collect survey data: Lessons learned from the IAB-SMART Study. *Survey Methods: Insights from the Field.*

► Harari, G. M., Müller, S. R., Stachl, C., Wang, R., Wang, W., Bühner, M., Rentfrow, P. J., Campbell, A. T., & Gosling, S. D. (2020). Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *Journal of Personality and Social Psychology, 119*(1), 204-228.

References

▶ Höhne, J. K. & Schlosser, S. (2019). SurveyMotion: What can we learn from sensor data about respondents' completion and response behavior in mobile web surveys? *International Journal of Social Research Methodology, 22*(4), 379-391.

▶ Jäckle, A., Burton, J., Couper, M. P., & Lessof, C. (2019). Participation in 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*(1), 23-44.

▶ Keusch, F., Bähr, S., Haas, G.-C., Kreuter, F., & Trappmann, M. (2020). Coverage error in data collection combining mobile surveys with passive measurement using apps: Data from a German national survey. *Sociological Methods & Research*, 1-38.

▶ Keusch, F., Wenz, A., & Conrad, F. (2022). Do you have your smartphone with you? Behavioral barriers for measuring everyday activities with smartphone sensors. *Computers in Human Behavior, 127*, 107054.

References

- Kreuter, F., Haas, G.-C., Keusch, F., Bähr, S., & Trappmann, M. (2020). Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review, 38*(5), 533-549.

- Lathia, N., Sandstrom, G. M., Mascolo, C., & Rentfrow, P. J. (2017). Happier people live more active lives: Using smartphones to link happiness and physical activity. *PLOS ONE, 12*(1), e0160589.

- MacKerron, G. & Mourato, S. (2013). Happiness is greater in natural environments. *Global Environmental Change, 23*, 992-1000.

- Scharkow, M. (2016). The accuracy of self-reported Internet use—A validation study using client log data. *Communication Methods and Measures, 10*(1), 13-27.

- Scherpenzeel, A. (2017). Mixing online panel data collection with innovative methods. In S. Eifler & F. Faulbaum (Eds.), *Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung* (pp. 27-49). Springer.

▶ Struminskaya, B., Lugtig, P., Keusch, F., & Höhne, J. K. (2020). Augmenting surveys with data from sensors and apps: Opportunities and challenges. *Social Science Computer Review*, 1-13.

▶ Sugie, N. (2018). Utilizing smartphones to study disadvantaged and hard-to-reach groups. *Sociological Methods & Research, 47*(3), 458-491.

▶ Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., & Campbell, A. T. (2014). StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 3-14).

▶ Wenz, A., Jäckle, A., & Couper, M. P. (2019). Willingness to use mobile technologies for data collection in a probability household panel. *Survey Research Methods, 13*(1), 1-22.

Photo Credit

- ▶ Slides 13 & 14: `https://unsplash.com/photos/EhQkcQ7k0pM`
- ▶ Slide 15: `https://unsplash.com/photos/rEn-AdBr3Ig`
- ▶ Slide 16: `https://unsplash.com/photos/vkkl0SeTnf4`
- ▶ Slide 18: `https://unsplash.com/photos/obZx1LjKKjc`
- ▶ Slide 19: `https://unsplash.com/photos/iD6mmn89YX4`