Emergent Issues in the Combined Collection of Self-Reports and Passively Measured Data Using Smartphones

Frederick Conrad, *University of Michigan*
Florian Keusch, *University of Mannheim*

New Measurement Possibilities

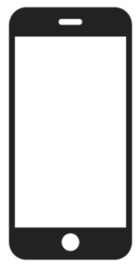
- Integration of smartphones into daily life enables in situ measurement -- at scale
  - Smartphones are an "extension of the body" so measurement is possible almost anytime, anywhere
  - Smartphones are nearly ubiquitous and users are eager to carry them so no additional devices and no special instructions from researchers required
  - Possible to collect both self-reports and to passively measure behaviors
- Self-reports: brief questionnaires, conditional on
  - Time: EMA (ESM)
  - Location: Geofencing
  - Behavior: Caller ID, app use, activity data
- Passively measured data from native sensors, project-specific apps:
  - This is Big Data, i.e., high volume (compared to self reports), high velocity, and without fixed structure or format

2

Native Smartphone Sensors

Proximity sensor
Bluetooth NFC

Compass
Barometer



Air humidity sensor

Camera*

GPS
Cellular network
Wi-Fi

Light sensor
Microphone

Fingerprint sensor

Thermometer

Accelerometer Pedometer
Gyroscope

*can be used to measure heart rate (Android/Samsung), linear distance (iPhone Measure app), as well as for photography

3

Goals and Rationale of Current Project

- Rapidly evolving paradigm
  - Research that jointly collects self-report and passive measures scattered across disciplines
  - Study designs vary but beginning to converge
- Sufficient number of studies now completed that possible to take stock of early successes and challenges, bringing research together in one place
- Makes it possible to
  - See emerging trends and connections
  - Understand the type of research questions that can be addressed with this approach
  - Identify methodological issues with the approach
- We will present insights we have gained from examining studies
  - We won't be presenting our original research

4

Benefits of Smartphones for Social Research

- Temporal granularity of passive measurement
  - Allows researchers to measure and evaluate moment-to-moment changes in behavior
  - Without increasing burden on participants
- Passive measures do not suffer from forgetting or other psychological processes that can lead to over- or underestimates in self-reports
- Combining passive measurement with self-reports provides a richer picture of social phenomena than either type of data alone
- Passively collecting data from a sample survey (not necessarily with self-report) can help overcome some limitations of Big Data
  - Design: Researcher can pre-specify field period, participant's characteristics, particular sensors used; introduces some control over the process
  - Representative sample: possible, in principle, to produce population estimates based on passively collected data

5

Errors of Representation

- Even where smartphone penetration is high, there is real risk of **coverage error** because those who do not own the devices are likely to differ profoundly from owners on many variables
  - Researchers have provided smartphones to participants in studies of the general population (Scherpenzeel 2017), the elderly (York Cornwell & Cagney 2017), chronically ill (Goodspeed et al. 2018), men recently released from prison (Sugie 2016), and students (Wang et al., 2014)
- Even if smartphone penetration were universal, the studies from representative samples show relatively **low participation rates** (16 - 22%)
  - Reasons for non-participation: intrusiveness, privacy, ability to use only limited features (Wenz et al., 2017; Keusch et al., in press)
  - In studies that collect both self-report and sensor data from the same smartphone users, the researcher is in effect asking users to make a *dual participation decision*

6

Measurement Error

- Passive measurement is not perfectly accurate

1. Sensors might be inaccurate
   - Two different devices (e.g., smartphone vs. GPS watch), two brands of smartphones (e.g., iPhone vs. Samsung), two generations of the same smartphone (e.g., iPhone 6 vs. iPhone X), or even two versions of the same operating system (e.g., Android 6.0 vs. Android 9.0) can produce different measures (e.g., Goodspeed et al. 2018; Höchsmann et al. 2018)
2. Inference from raw sensor data to target behavior (e.g., from phone's movement to "steps" to "walking") may lead to misclassification (invalidity)
   - e.g., sleep may be defined by darkness, silence and lack of movement detected by sensors (Wang et al., 2014) could be due to phone left in bag at home
3. Non-compliance
   - Unintentional (e.g., forget to turn phone back on after movie) -- probably increasing variance
   - Deliberate (e.g., app turned off to hide undesirable behaviors or locations) -- producing bias
   - But evidence, so far, is that compliance is high (Kreuter et al., in press; Sugie, 2018)

7

Cost

- Passive data collection requires little effort for either the participant or the researcher and is in some ways inexpensive, i.e., few variable costs
- But costs for…
  - App development potentially for different operating systems -- currently not many off-the-shelf platforms (e.g., Apple Research Kit, ResearchStack for Android)
  - IT infrastructure (i.e., storage capacity and skilled support staff) for big datasets
  - Data preparation, analysis, and reporting capability (e.g., data wrangling, machine learning, data linkage, visualization, etc.)
- Taken together, these factors can lead to a far more expensive project than might be evident to the researcher using these methods for the first time

8

Review of Relevant Studies

- Inclusion criteria:
  - Passive measurement and self-report on smartphone within participants
- Included studies*
  1. Green et al. (2016): In the Moment Travel Study (rMove)
  2. Goodspeed et al. (2018):
  3. Kreuter et al. (under review): IAB-SMART
  4. MacKerron & Mourato (2013): Mappiness
  5. Scherpenzeel (2017): Smartphone Time Use Study
  6. Scherpenzeel (2017) & Geurs et al. (2015): Mobile Mobility Study (MoveSmarter)
  7. Sugie (2018): Newark Smartphone Reentry Project
  8. York Cornwell & Cagney (2017): Real Time Neighborhoods and Social Life Study
  9. Wang et al. (2014): StudentLife

*As far as we know this exhausts set of relevant studies but we could be missing some

Study Domain*

- Mobility (2)
- Relationship between place, mobility, and health
- Employment, unemployment, and poverty
- Happiness across physical settings
- Time use
- Social reintegration of parolees
- Effect of locations, activities, and experiences on health and well-being
- Student well-being over the course of an academic term

*Although all studies conducted in an application domain, they have the character of feasibility studies

Types of Data Collected Across the Studies

- Sensors measure objective phenomena
  - Location, movements, and trips (9)
  - Interaction by phone, by text, and face-to-face (4)
  - Type of activities (2)
  - Smartphone use (1)
  - Characteristics of social network (1)
  - Proximity to others (1)
  - Sleep (1)
- Self-reports measure subjective and objective phenomena
  - Affect/mood/stress/fatigue/anger... (5)
  - Trip details (e.g., purpose, travelers) (2)
  - Activity categories (2)
  - Job search (2)
  - Health status (2)
  - Demographics (2)
  - Personality -- Big 5 (1)
  - Location (1)
  - Smartphone use (1)
  - Social interaction (1)

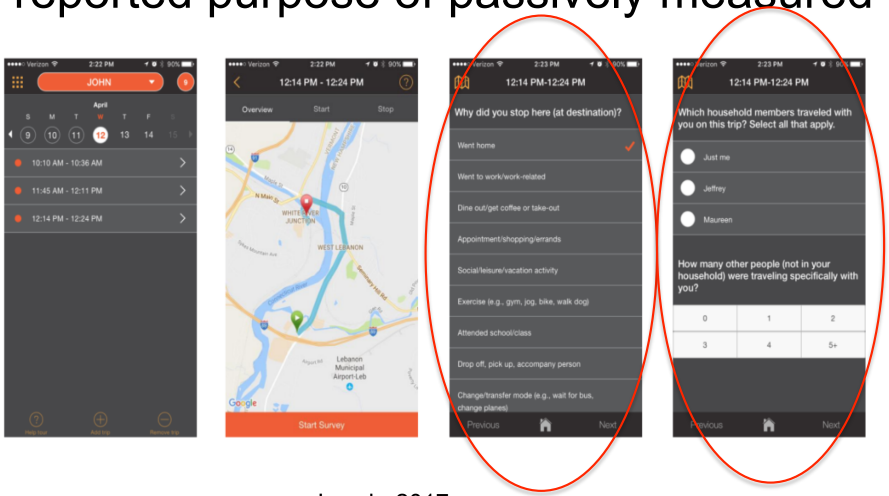
11

Synergy in Joint Use of Self-Reports and Sensor Data

- Self-reports can be used to <u>verify</u> sensor measurement
  - e.g., participants confirm or edit locations and modes of transport (EMA) for passively measured trips (Scherpenzeel, 2017)
- Self-reports can provide <u>context</u> for sensor data, promoting their interpretation
  - e.g., EMA questions ask about <u>purpose</u> of passively measured trip (Green, 2016; Lynch 2017)

12

Self-reported purpose of passively measured trip



Lynch, 2017

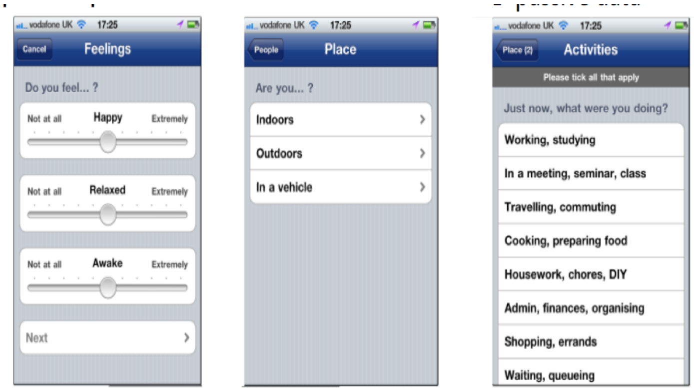
13

Synergy in Joint Use of Self-Reports and Sensor Data

- Self-reports can be used to verify sensor measurement
  - e.g., participants confirm or edit locations and modes of transport (EMA) for passively measured trips (Scherpenzeel, 2017)
- Self-reports can provide context for sensor data, promoting their interpretation
  - e.g., EMA questions ask about purpose of passively measured trip (Green, 2016)
- and sensor data can provide context for self-reports
  - e.g., passively measured locations (e.g., green/natural habitat, urban) help explain subjective well-being measured by EMA: participants happier outdoors and when sunny (MacKerron and Mourato, 2013).

14

Self-reported mood is explained by passively collected location data



MacKerron & Murato, 2013

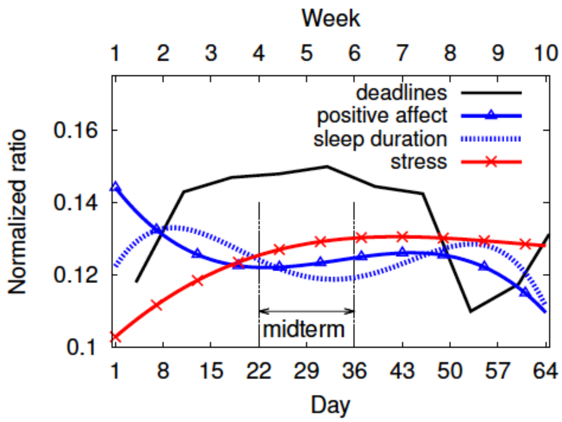
15

Synergy in Joint Use of Self-Reports and Sensor Data

- Self-reports can be used to <u>verify</u> sensor measurement
  - e.g., participants confirm or edit locations and modes of transport (EMA) for passively measured trips (Scherpenzeel, 2017)
- Self-reports can provide <u>context</u> for sensor data, promoting their interpretation
  - e.g., EMA questions ask about <u>purpose</u> of passively measured trip (Green, 2016)
- and sensor data can provide <u>context</u> for self-reports
  - e.g., passively measured locations (e.g., green/natural habitat, urban) help explain subjective well-being measured by EMA: participants happier outdoors and when sunny (MacKerron and Mourato, 2013).
- Self-reports and sensor data can be <u>correlated</u>; neither is primary data source
  - Self-reported stress (EMA) increases as passively measured sleep is reduced, related students' academic deadlines (e.g., Wang et al., 2014)

16

Correlation between Sensor and Self-report data



Depression* and Sensing Data

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

*measured with Patient Health Questionnaire at start and end of semester, i.e., not strictly *in situ*

Wang et al, 2014

17

Synergy in Joint Use of Self-Reports and Sensor Data

- Self-reports can be used to verify sensor measurement
  - e.g., participants confirm or edit locations and modes of transport (EMA) for passively measured trips (Scherpenzeel, 2017)
- Self-reports can provide context for sensor data, promoting their interpretation
  - e.g., EMA questions ask about purpose of passively measured trip (Green, 2016)
- and sensor data can provide context for self-reports
  - e.g., passively measured locations (e.g., green/natural habitat, urban) help explain subjective well-being measured by EMA: participants happier outdoors and when sunny (MacKerron and Mourato, 2013).
- Self-reports and sensor data can be correlated; neither is primary data source
  - Self-reported stress (EMA) increases as passively measured sleep is reduced, related students' academic deadlines (e.g., Wang et al., 2014)
- Passively detected location can trigger self-report measurement
  - Kreuter et al. (under review) asked questions about job seeking experience when passive location measurement indicated at local employment agency ("geofencing")

18

Frequency of Passive Measurement

- Passive measurement inherently longitudinal but exact frequency of measurement is up to researcher

1. Situational measurement
   - initiated when certain conditions (geographical, temporal, social/proximal, etc.) are met
   - For example, MacKerron and Mourato's mappiness app collects participant's geololation when participants pinged with EMA questions (twice daily at random times)
2. High frequency, discrete
   - More often than self-reports can reasonably be collected but data collection not *always-on*
   - For example, Kreuter et al. (under review) measured geolocation every 30 minutes over six months; EMA questionnaire pushed to participants every few days
3. Continuous
   - Capturing fine-grained behavior such as movement requires continuous measurement
   - For example, Wang et al. (2014) used the accelerometer to infer activity states (stationary, walking, running, driving, cycling), as well as microphone and light sensor to measure sleep
   - Battery life and storage considerations

19

Implications for Privacy

- High frequency measurement, especially which participant does not control, likely to seem intrusive, potentially reducing consent or completion of study
- Several steps have been used to reduce potential participants' concerns
  - Explain why passively collected data are required to meet project's scientific goals and how participants' data and identities will be protected
  - Allow participants to turn off tracking temporarily
  - Limit sensors' abilities, e.g., Wang et al. (2014) used "privacy-sensitive audio and conversation classifiers" to determine participant was "around conversation" but classifiers could not reconstruct the content of speech or identify individual speaker
  - When using apps that track voice calls and text messages, record only first name of the other person (Sugie, 2018) or encrypt phone number (Kreuter et al., in press)

20

Discussion/Conclusion

- Joint collection of self-report and sensor data on smartphones in early stages
- But several themes are emerging:
  1. Combining these two types of data really can provide insights not likely using either type of data alone
  2. Most widely used so far are passively measured location, mobility, and activity (e.g., GPS, accelerometer) combined with time-contingent EMA
  3. Using multiple sensors (e.g., microphone, light sensor, gyroscope) allows researchers to infer behaviors (e.g., sleep) for which no one sensor exists
  4. Passively measuring behavior can reduce need to collect some self-reports, e.g., location, sleep duration, amount of social contact, steps
     - These passive measures serve, at least, as a good proxy for self-reports
     - But sometimes a better measure
     - With less burden

21

Discussion/Conclusion (2)

- Approach is still being evaluated methodologically, i.e., not yet widely used to address substantive research questions
  - Focus is still on determining what is feasible and assessing the quality of the data
- Much evaluation still needed, e.g.,
  - Non-participation bias: are those who participate systematically different on measured behaviors than non-participants (e.g., are participants in mobility study more active than non-participants)?
  - Misclassification of sensor data (e.g., mistaking data from phone left at home in a bag as evidence participant is asleep): how common are these errors and how can they be reduced?
- The approach has inherent limits
  - Many behaviors cannot be captured via passive measurement on smartphones (e.g., illicit drug use, dietary intake)
- May work especially well in longitudinal panel research because of trust and familiarity established between researcher and participant over time

22

Discussion/Conclusion (3)

- Reactivity?
- Can <u>attitudes</u> be measured with some combination of sensors?
- Impact on social and behavioral research likely increase as new and better sensing technologies emerge and creative ways to use sensors, especially in conjunction with with self-reports, are developed

23

Thank you!

24