Smartphone devices data in health research: opportunities and challenges

Marta Karas

Wearable and Implantable Technology (WIT) group meeting

Mar 4, 2022

Overview

- Smartphone devices data in health research
- 2 application examples
- Challenges in data collection
- Opportunities in data collection
- Challenges in data analysis
- Opportunities in data analysis

Follows and borrows from:

Onnela, JP. *Opportunities and challenges in the collection and analysis of digital phenotyping data*. Neuropsychopharmacol. 46, 45–54 (2021). https://doi.org/10.1038/s41386-020-0771-3

Smartphone devices data

Smartphones

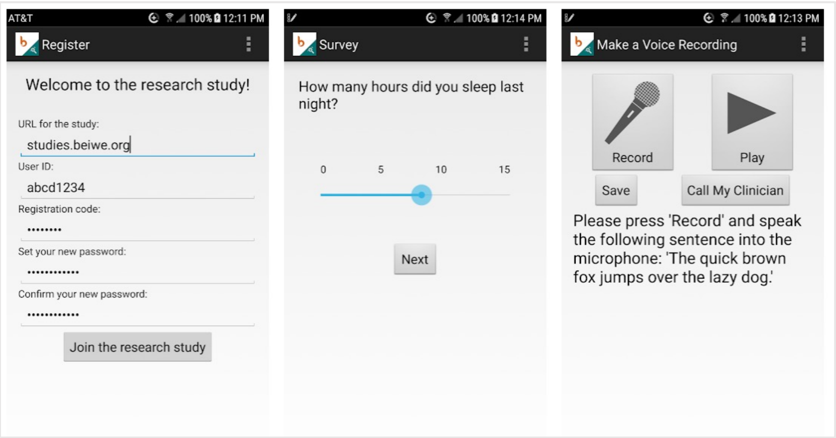
- Broadly adopted and used
- Allow data collection using one's existing personal device

"Active data"

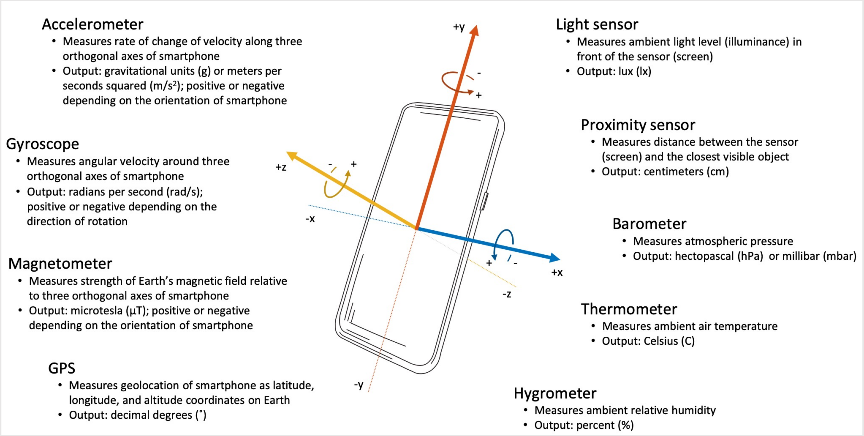
- Requires the user to actively participate in a data collection activity
- Taking surveys, contributing audio diary entries, carry out cognitive assessments

"Passive data"

- Can be generated by the device passively, pose no burden on the participant
- From smartphone sensors and smartphone logs (e.g., communication logs, screen activity logs)



Smartphone app, a part of the Beiwe research platform (Harvard University, Onnela Lab) to collect smartphone sensor and usage data in clinical and non-clinical studies



Straczkiewicz et al. (2021). Overview of standard smartphone sensors

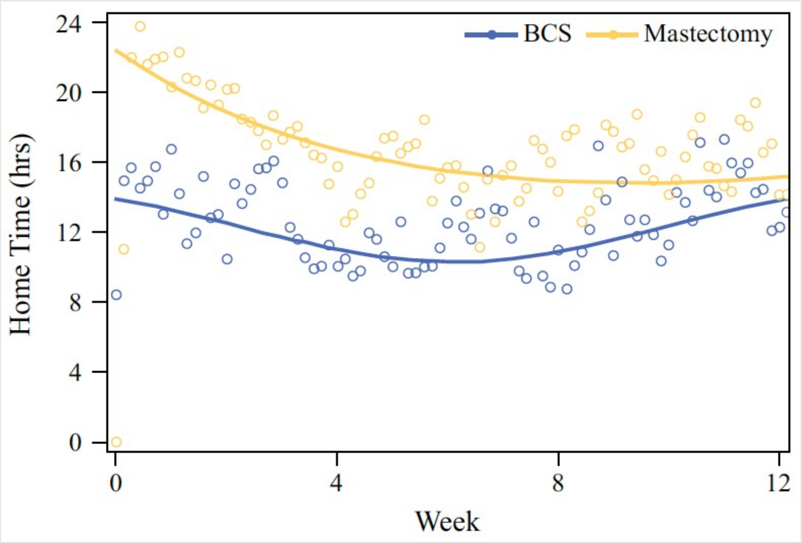
Example: using GPS to quantify PA differences in recovery after BCS and mastectomy

GPS data-derived measures

- Distance traveled each day -- may reflect changes in gross mobility, pain, and fatigue
- Time spent at home -- may be associated with loneliness, community engagement, and amount of social interactions

Panda et al. (2021)

- N=31 patients (16 BCS, 15 mastectomy), no signif. differences in demographics, home time, distance travelled at baseline
- Collected GPS data 1 week preoperatively and 6 months postoperatively
- Through 12 weeks postoperatively, mastectomy patients spent more time at home and traveled shorter distances



Panda et al. (2021). Trends in postoperative daily home time in first 12 weeks after surgery as derived from smartphone GPS data.

Example: predicting schizophrenia relapse with active and passive data

Smartphone-collected survey data

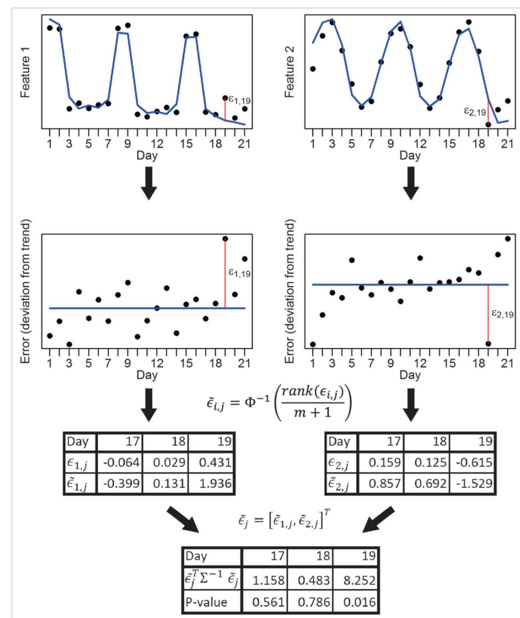
- Offer insights into present life experiences, psychiatric symptoms
- Frequency and timing must be considered

Barnett et al. (2018)

- N=17 patients with schizophrenia in active treatment, followed for 3 months
- Beiwe app used to collect symptom surveys and passive data
- Tested for anomaly occurrence in multiple streams data, investigated the rate at which significant anomalies occur
- Rate of anomalies detected in the 2 weeks prior to relapse was 71% higher than the rate of anomalies further away

Survey question categories	Mobility features	Sociability features
1. Depression	1. Time spent at home	1. Number of outgoing texts
2. Sleep quality	2. Distance traveled	2. Total outgoing text length
3. Psychosis	3. Radius of gyration	3. Texting out-degree
4. Warning symptoms scale	4. Maximum diameter	4. Number of incoming texts
5. Taking medication	5. Maximum distance from home	5. Total incoming text length
6. Anxiety	6. Number of significant locations	6. Texting in-degree
	7. Average flight length	7. Texting reciprocity
	8. Standard deviation of flight length	8. Texting responsiveness
	9. Average flight duration	9. Number of outgoing calls
	10. Standard deviation of flight duration	10. Total outgoing call duration
	11. Fraction of the day spent stationary	11. Call out-degree
	12. Significant location entropy	12. Number of incoming calls
	13. Minutes of GPS data missing	13. Total incoming call durations
	14. Physical circadian rhythm	14. Call in-degree
	15. Physical circadian rhythm stratified	15. Call reciprocity
		16. Call responsiveness

Barnett et al. (2018). Mobility and sociability features calculated each day for each patient.



Barnett et al. (2018). Multivariate time-series anomaly detection method based on Hotelling's T-Square test

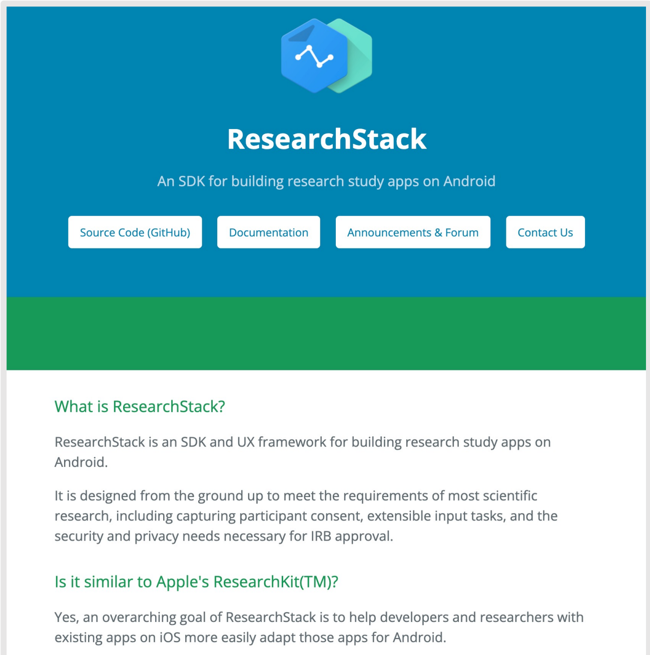
Challenges in data collection

- Data collection happens through a smartphone application
- **Active data: challenge of how to keep subjects engaged**
  - Use of financial incentives, provide feedback, use gamification
- **Passive data: access**
  - Originates from smartphone sensors (GPS, accelerometer) and smartphone logs (communication logs, screen activity logs)
  - Sensors: application has to request access to a sensor for that data stream to become available; no way to collect sensor data retrospectively
  - Logs: some smartphone logs are available and accessible for times that precede installation (Android communication logs)

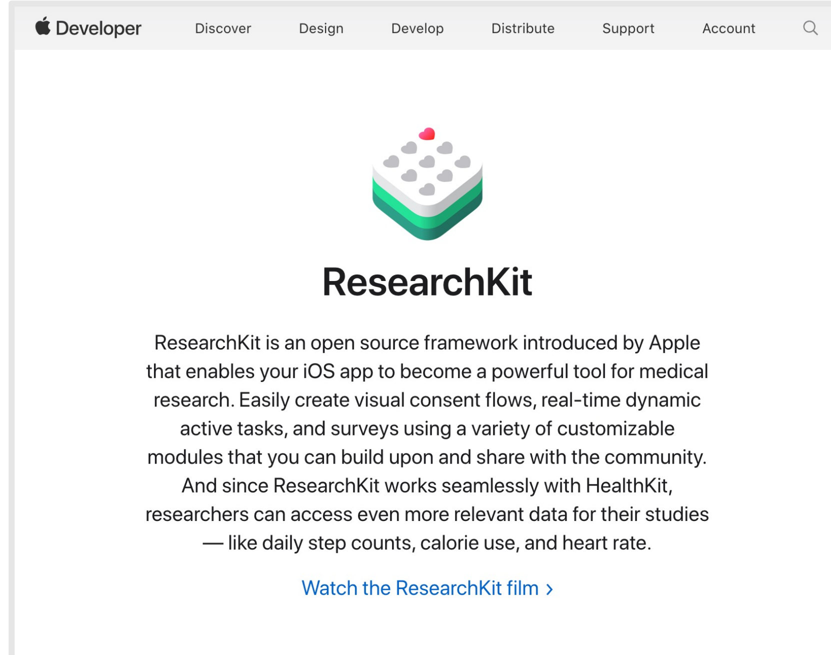
Challenges in data collection (cont.)

- **Passive data: continuous background collection**
  - Application is running in the foreground (is shown on the screen and user can interact with it) vs. in the background (user is not using application)
  - **Software development kits (SDKs)** for smartphones exist to facilitate building research study apps
    - Google's ResearchStack (Android),
    - Apple's ResearchKit, HealthKit, CareKit (all iOS)
  - Use of prepackaged software limits what type of data can be collected
    - ResearchKit does not support background sensor data collection
    - HealthKit supports background sensor data collection but only in a limited manner (up to up to 12 h of accelerometer data as of 2021)
    - Some of the algorithms are proprietary and are subject to change
  - => Continuous background data collection usually requires writing custom code

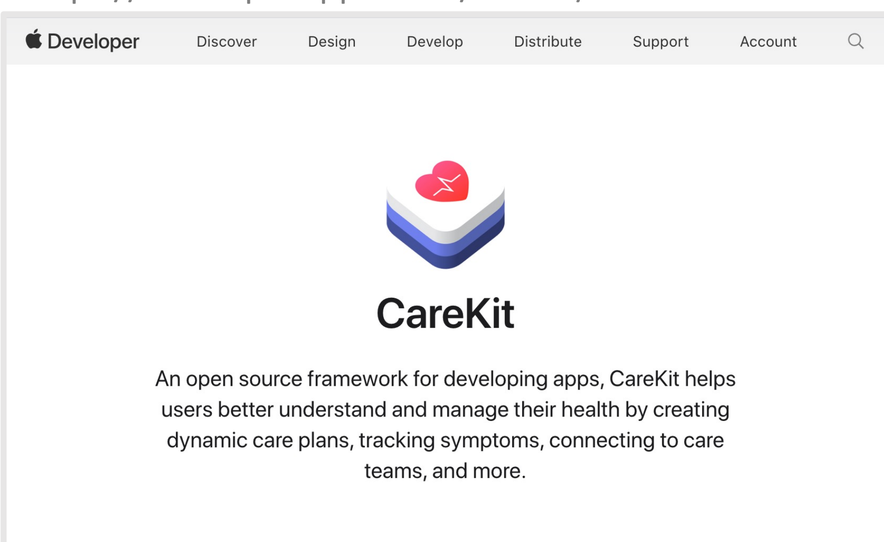
http://researchstack.org/



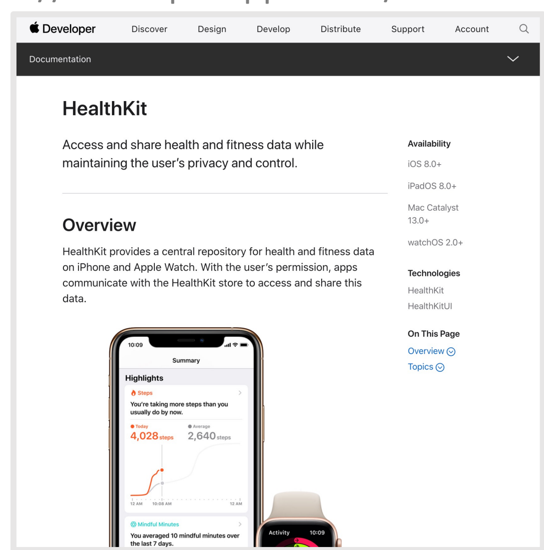
https://developer.apple.com/researchkit/



https://developer.apple.com/carekit/



https://developer.apple.com/documentation/healthkit



Challenges in data collection (cont.)

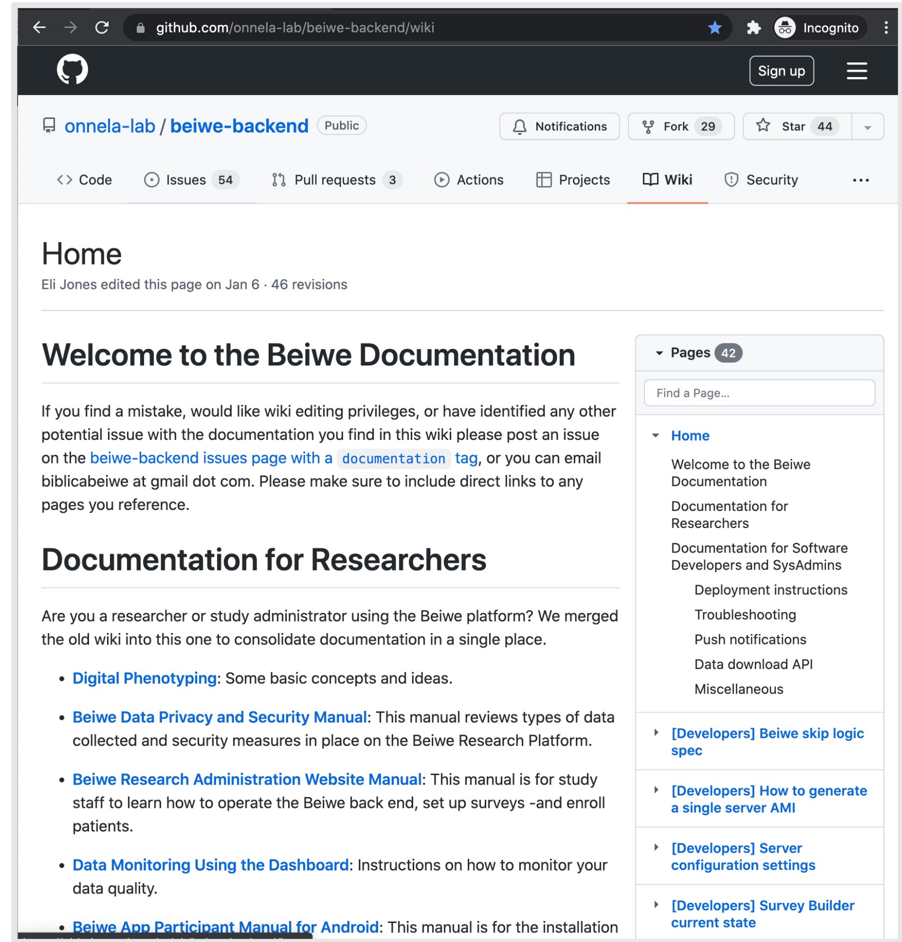
- **Challenges to privacy and data security**
  - Collected data must be **encrypted** at all times -- while stored on the phone awaiting upload, while in transit, for storage on the study server while at rest
  - Encryption keys pair may be used to decrypt data on server
  - Some data contain identifiers (phone numbers from communication logs on Android devices); may **anonymize** via hashing algorithms
  - Some data can indicate the location of a person's home (GPS data); may use the "noisy GPS" data stream which incorporates additive noise to the coordinates

Opportunities in data collection

- Existence of **open-source research platforms** that can be used to conduct smartphone-based monitoring studies
  - **Front-end**: smartphone application (supporting both Android and iOS
    - Collect various types of data: surveys and audio diary entries, GPS, raw accelerometer data, anonymized call and text logs
  - **Back-end**: storage / computing infrastructure (e.g., AWS)
    - Store the collected data, support study monitoring and data analysis activities
  - Enable **customization** in active and passive data collection
- Enables to to study:
  - Behavioral patterns, social interactions, mobility, gross motor activity, and speech production

https://www.beiwe.org/

https://github.com/onnela-lab/beiwe-backend/wiki



Opportunities in data collection (cont.)

- Instrument for **experience sampling methods (ESM)** and **ecological momentary assessment (EMA)**
  - ESM/EMA questions pertain to experiences from present moment / recent interval
  - Memory biases are expected to be minimal
- Especially important for **psychiatric disorders studies**
  - Koster et al. (2010): Current mood is likely to influence the type of information that is recalled
  - Myin-Germeys et al. (2016): Retrospective reports of extreme mood changes, over the previous month and even over the preceding week, were largely unrelated to reports obtained when considering the present moment
  - Solhan et al. (2009): Retrospective recall of average levels of mood or symptoms might be also more difficult than considering the present moment, particularly for individuals with psychiatric diagnoses

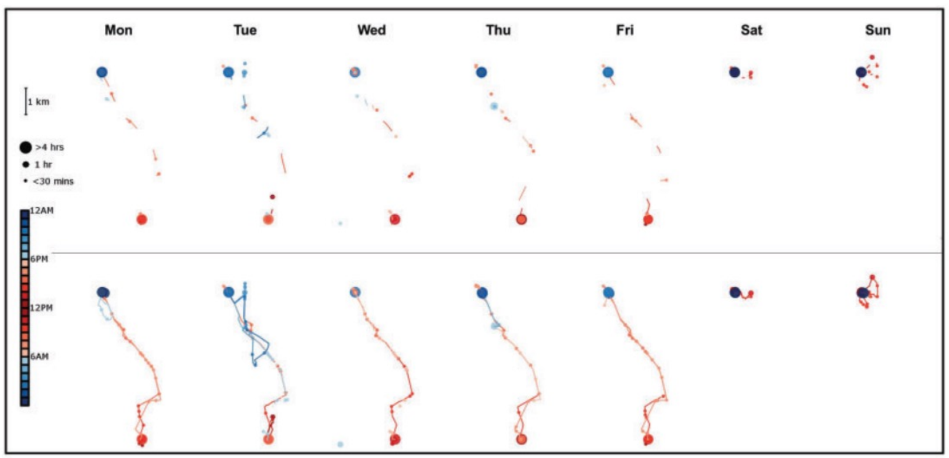
Challenges in data analysis

- A blessing and a curse: "*Smartphones are personal devices that are used frequently by many people, over long periods of time, in a myriad of ways*"
- **Accounting for different ways how people use their phones**
  - Orientation of the device changes
  - Some individuals turn their phone off for the night
  - Whether the phone is at person at different times
- **Missing data**
  - Some sensors need to be sampled at different cycles for different reasons, so some missingness is expected by design (e.g., on-cycles and off-cycles of GPS or raw accelerometer data sampling to avoid draining the battery)
- Propagating the uncertainties involved at different stages of the process in data analysis

Passive data: imputation of GPS data

GPS data collection

- GPS is typically active only for a small portion of the time (e.g., < 10%)
- Fixed-length on and off cycles
- => missing data problem



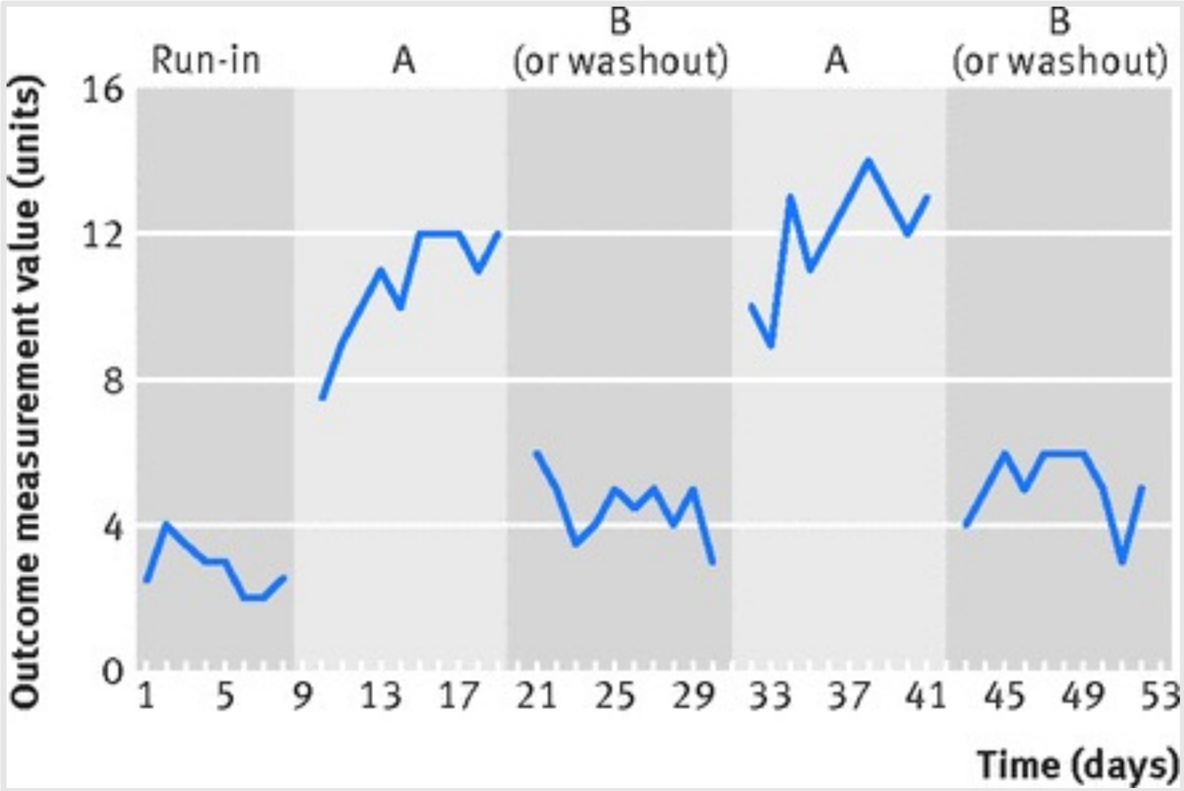
Barnett et al. (2020). A person’s daily trajectories over the course of a week. Top: when GPS data 2 min + 10 min on and off cycles. Bottom: when GPS collected continuously.

Barnett at al. (2020): impute the missing mobility traces based on weighted resampling of observed data

- Map location coordinates (ϕ, λ) into $[X, Y]$ plane
- Transform into sequence of: flights (linear move), pauses (no move), missing data: $\Delta_j = (\Delta_j^x, \Delta_j^y, \Delta_j^t)$
- Sample displacements Δ_j for missing periods from distribution of observed flights and pauses, conditional on the time and location

Opportunities in data analysis

- **Moving to real time (or close to real-time) analyses** from retrospective analyses carried out after data analysis has been completed
    - Actionable for interventions based on some changes or anomalies in data
    - Especially well-suited for interventions in high-risk populations/setups studies (psychosis, suicidality)
    - Leveraging online algorithms that process data in a piece-by-piece fashion and generates output along the way vs. offline algorithms that need the entire dataset to output an answer
- **Conducting N-of-1 (single subject) trials**
    - Consist of two or more treatments for a single individual where the sequence of the treatments may or may not be randomized
        - For example, a sequence "ABAB" of treatments "A" and "B" constitutes a four-period crossover design
    - Multiple measurements are needed (compared with traditional randomized clinical trials where the primary end point might be measured only twice)



BMJ 2015;350:h1793:
N-of-1 trial pictorial showing participant's progress and outcomes through an individual trial.

Summary

1. Smartphone-based data allow for remote monitoring using already owned device
2. Active data (surveys, voice recordings) and passive data (GPS, communication logs) increasingly used in clinical and non-clinical research

References

1. Barnett, I., Torous, J., Staples, P., Sandoval, L., Keshavan, M., & Onnela, J.-P. (2018). Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology (New York, N.Y.), 43(8), 1660–1666. https://doi.org/10.1038/s41386-018-0030-z

2. Barnett, I. & Onnela, J.-P. (2020). Inferring mobility measures from GPS traces with missing data. Biostatistics (Oxford, England), 21(2), E98–E112. https://doi.org/10.1093/biostatistics/kxy059

3. Koster, E.H.W., De Raedt, R., Leyman, L., & De Lissnyder, E. (2010). Mood-congruent attention and memory bias in dysphoria: Exploring the coherence among information-processing biases. Behaviour Research and Therapy, 48(3), 219–225. https://doi.org/10.1016/j.brat.2009.11.004

4. Myin-Germeys, I., Klippel, A., Steinhart, H., & Reininghaus, U. (2016). Ecological momentary interventions in psychiatry. Current Opinion in Psychiatry, 29(4), 258–263. https://doi.org/10.1097/YCO.0000000000000255

5. Onnela, J.-P. (2021). Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacology (New York, N.Y.), 46(1), 45–54. https://doi.org/10.1038/s41386-020-0771-3

6. Panda, N., Solsky, I., Hawrusik, B., Liu, G., Reeder, H., Lipsitz, S., Desai, E. V., Lowery, K. W., Miller, K., Gadd, M. A., Lubitz, C. C., Smith, B. L., Specht, M., Onnela, J.-P., & Haynes, A. B. (2020). Smartphone Global Positioning System (GPS) Data Enhances Recovery Assessment After Breast Cancer Surgery. Annals of Surgical Oncology, 28(2), 985–994. https://doi.org/10.1245/s10434-020-09004-5

7. Solhan, M.B., Trull, T. J., Jahng, S., & Wood, P. K. (2009). Clinical Assessment of Affective Instability. Psychological Assessment, 21(3), 425–436. https://doi.org/10.1037/a0016869

8. Straczkiewicz, M., James, P., & Onnela, J.-P. (2021). A systematic review of smartphone-based human activity recognition methods for health research. NPJ Digital Medicine, 4(1), 148–148. https://doi.org/10.1038/s41746-021-00514-4

Thank you!

Marta Karas

mkaras@hsph.harvard.edu
https://martakarass.github.io/
@martakarass