# Digital Phenotyping

Jukka-Pekka "JP" Onnela

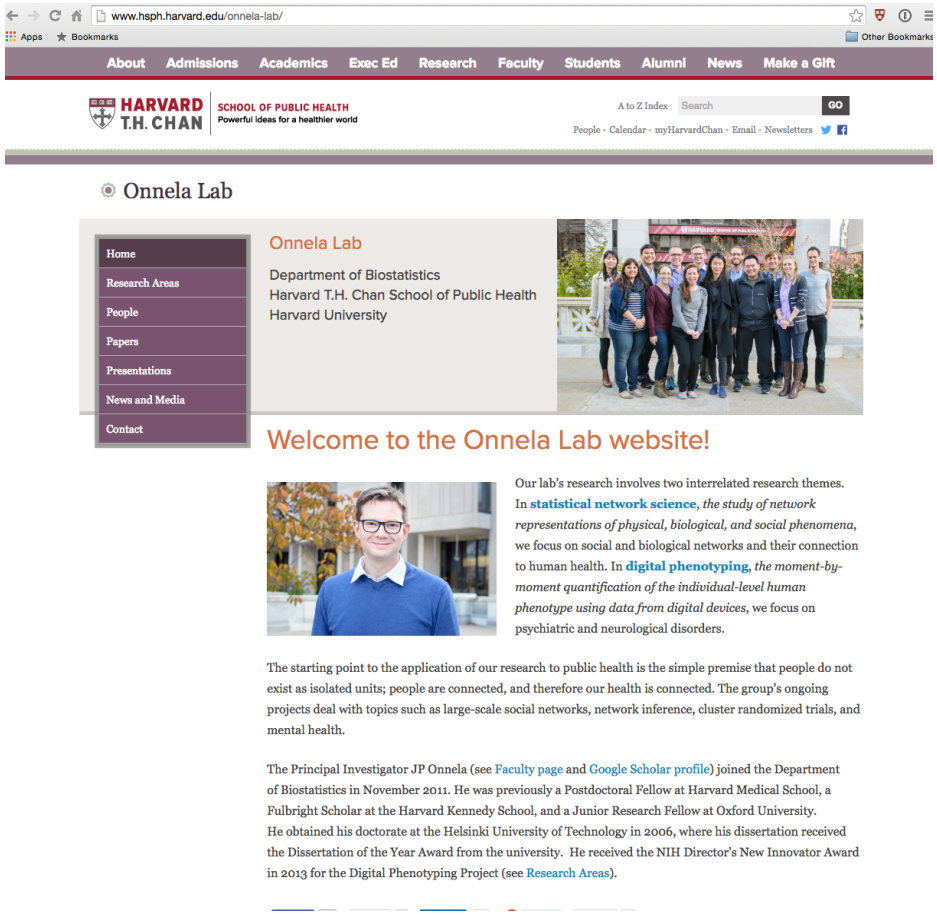
Department of Biostatistics

Harvard T. H. Chan School of Public Health

July 20, 2016

# INTRODUCTION

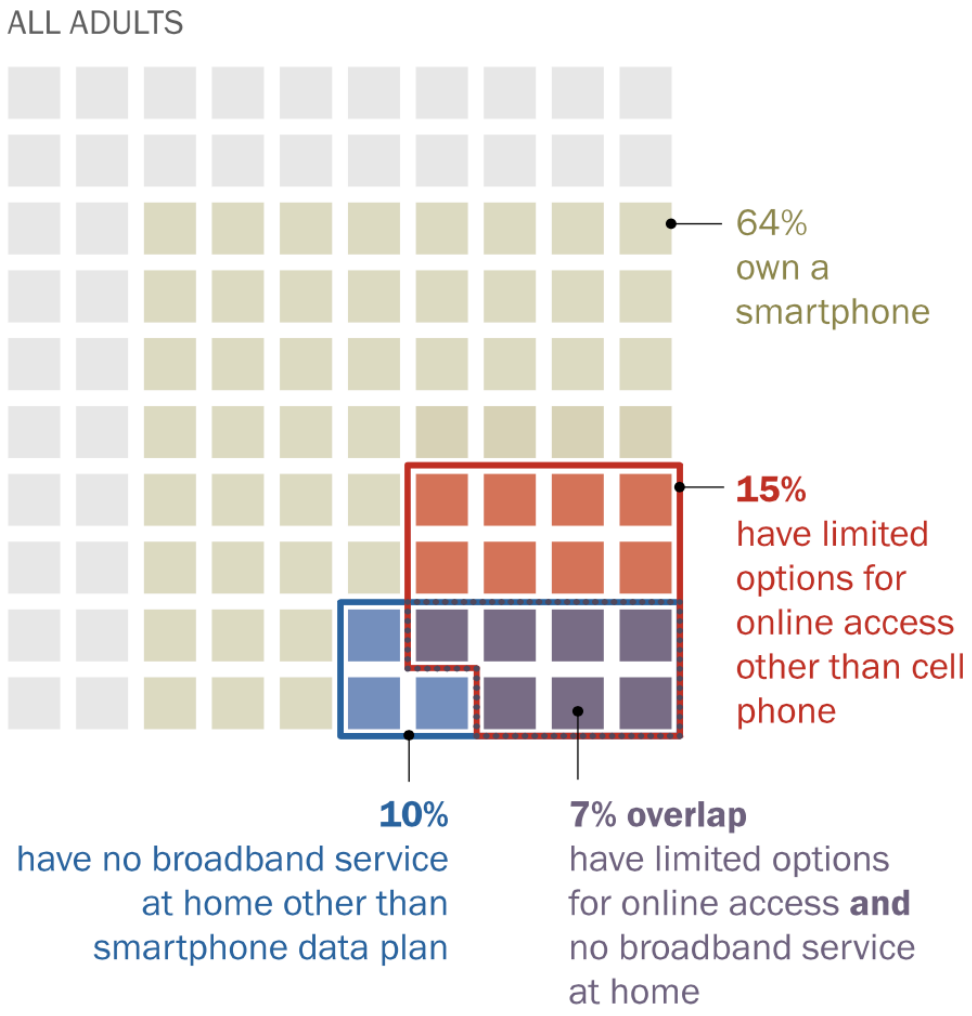
- Onnela Lab
  - **Statistical network science**
  - **Digital phenotyping**




**www.hsph.harvard.edu/onnela-lab/**

# SMARTPHONES

- **64% of American adults own a smartphone in 2015**
- **Up from 35% in 2011**




Pew Research Center American Trends Panel survey, October 3-27 2014.

U.S. Smartphone Use in 2015, Pew Research Center, 2015

http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/

# DIGITAL PHENOTYPING

- **Digital Phenotyping Project** (funded by 2013 NIH Director's New Innovator Award):
    1. Develop **customizable, scalable, open, research platform** for collecting smartphone data
    2. Develop **statistical methods** for analyzing and modeling the data

- Digital phenotyping
    - Definition: "Moment-by-moment quantification of the individual-level human phenotype *in situ* using data from personal digital devices"
    - Goal: Capture the lived experiences of subjects, and their interactions with the surrounding world, with minimal interference
    - Active data & passive data

# DIGITAL PHENOTYPING

**Overall goal**

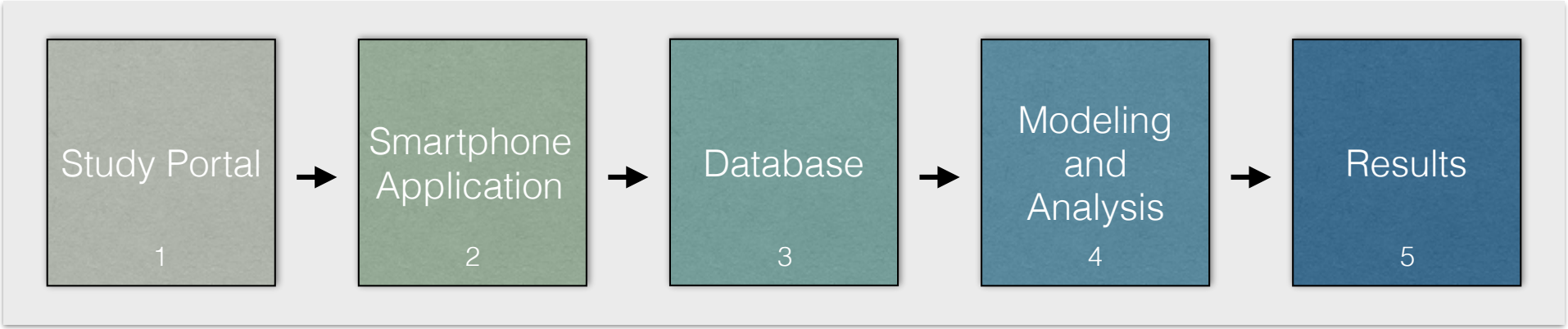
- **Tools → Data → Theory → Understanding**

**Scientific goals**

- Precise disease phenotypes for psychiatric, neurological, and surgical patients
- Classification of psychopathologies based on observable behavior
- Monitor response to treatment or intervention
- More efficient drug trials
- Exposure to environment
- Depth of genotype vs. depth of phenotype (GWAS)
- **Deep phenotyping (longitudinal)**

# DIGITAL PHENOTYPING

- Completed development of the **Beiwe research platform**:
  1. Manage study and data collection using the web-based study portal
  2. Download app(Android & iOS)
  3. Store hashed and encrypted data on cloud server
  4. Model and analyze data using open source Beiwe data tools
  5. Share "data protocol" and publish results




"New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research" by J Torous, MV Kiang, J Lorme, and JP Onnela. JMIR Mental Health 3(2):e16 (2016)

# DATA STREAMS

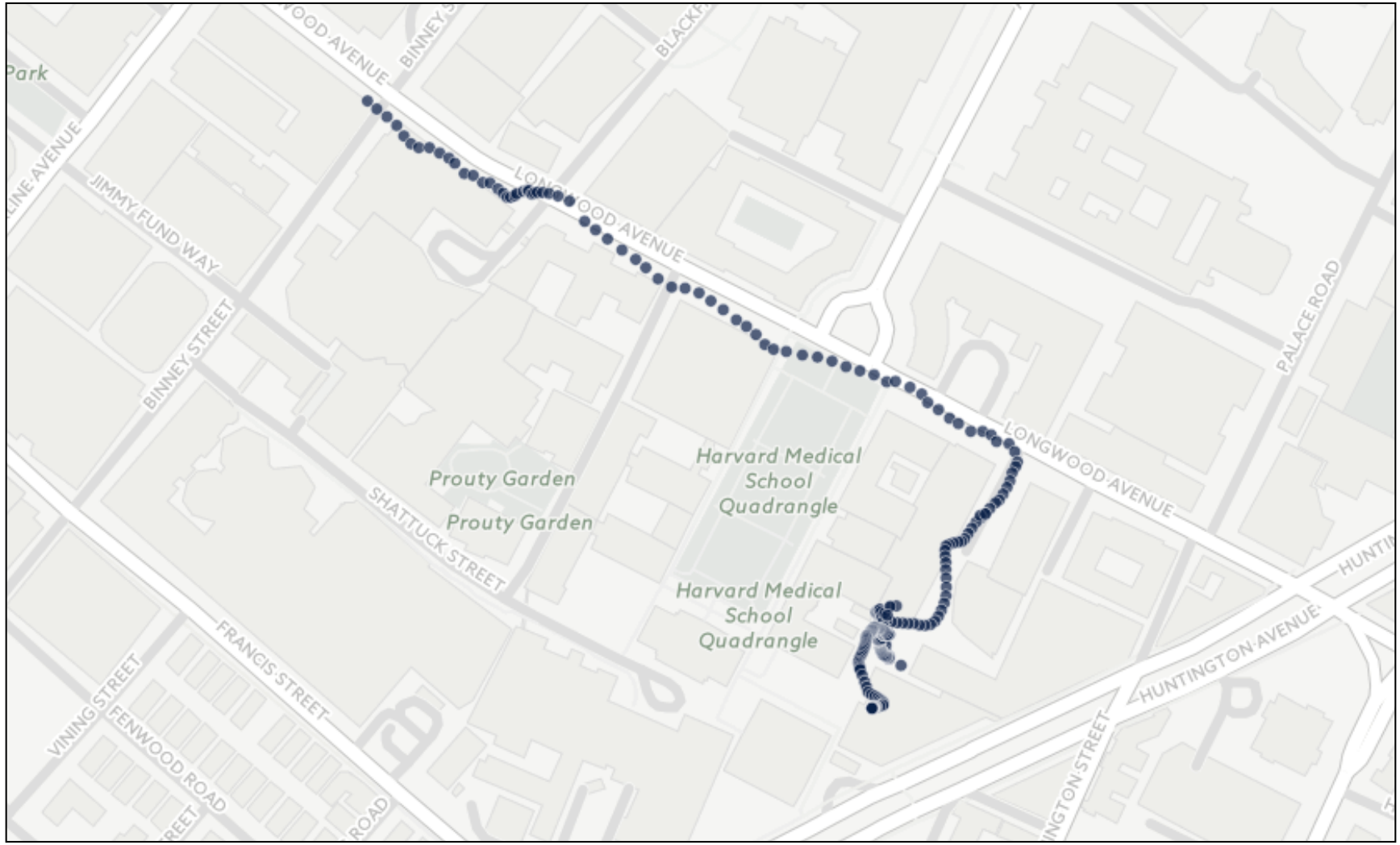
**Active data**

- Survey responses and metadata
- Voice / audio recordings
- Cognitive tests (under development)

**Passive data**

- GPS
- Accelerometer
- Phone and screen state
- WiFi routers
- Bluetooth devices
- Phone call logs
- Text message logs
- Magnetometer
- Proximity

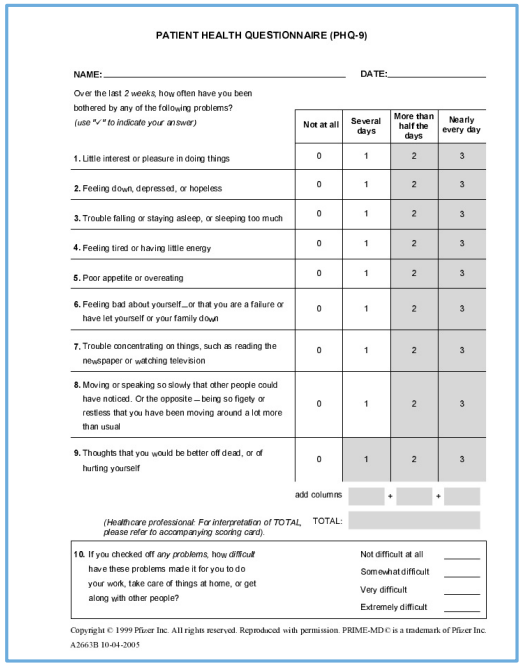
# BEIWE DATA: GPS




- Map: https://mkiang.cartodb.com/viz/c67b3202-2023-11e5-96ef-0e853d047bba/public_map
- Animation: http://cdb.io/1GvZefN

# PILOT STUDY

- Surveys (like PHQ-9) are the standard approach in mental health
- Completed during office visits
- Difficult to collect high frequency longitudinal data
- Rely on retrospective recollection and may be inaccurate
- Conformation to expectations or avoidance of responses
- This is an "in vitro" as opposed to an "in vivo" approach



**PATIENT HEALTH QUESTIONNAIRE (PHQ-9)**

NAME:_____ DATE:_____

Over the last 2 weeks, how often have you been bothered by any of the following problems?
*(use "✓" to indicate your answer)*

| | Not at all | Several days | More than half the days | Nearly every day |
|---|---|---|---|---|
| 1. Little interest or pleasure in doing things | 0 | 1 | 2 | 3 |
| 2. Feeling down, depressed, or hopeless | 0 | 1 | 2 | 3 |
| 3. Trouble falling or staying asleep, or sleeping too much | 0 | 1 | 2 | 3 |
| 4. Feeling tired or having little energy | 0 | 1 | 2 | 3 |
| 5. Poor appetite or overeating | 0 | 1 | 2 | 3 |
| 6. Feeling bad about yourself—or that you are a failure or have let yourself or your family down | 0 | 1 | 2 | 3 |
| 7. Trouble concentrating on things, such as reading the newspaper or watching television | 0 | 1 | 2 | 3 |
| 8. Moving or speaking so slowly that other people could have noticed. Or the opposite – being so figety or restless that you have been moving around a lot more than usual | 0 | 1 | 2 | 3 |
| 9. Thoughts that you would be better off dead, or of hurting yourself | 0 | 1 | 2 | 3 |
| add columns | | + | + | |
| *(Healthcare professional: For interpretation of TOTAL, please refer to accompanying scoring card).* | TOTAL: | | | |

10. If you checked off *any problems*, how *difficult* have these problems made it for you to do your work, take care of things at home, or get along with other people?

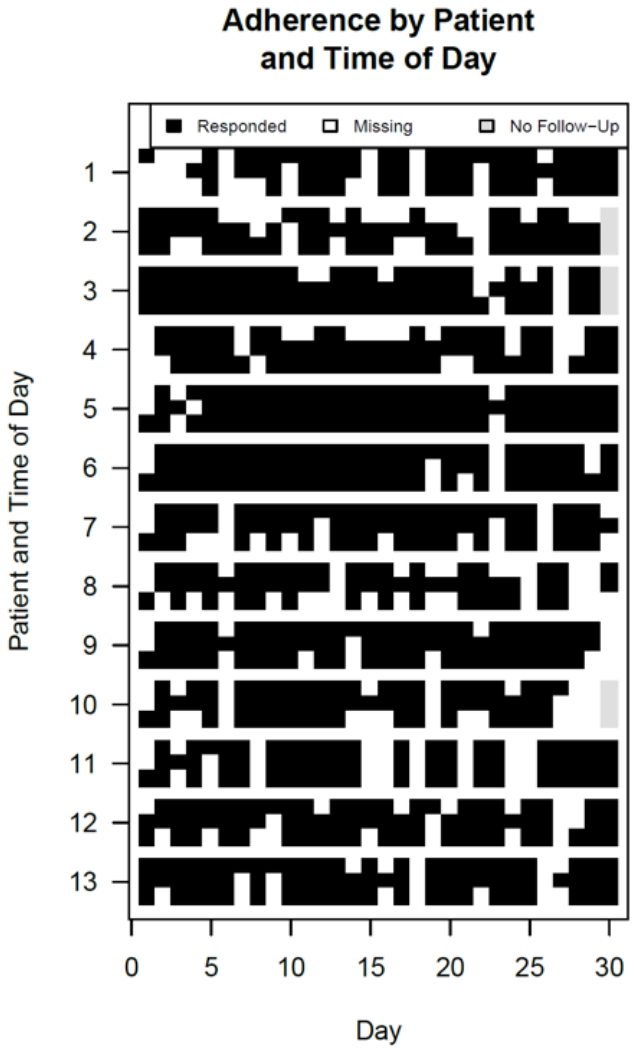
- Not difficult at all _____
- Somewhat difficult _____
- Very difficult _____
- Extremely difficult _____

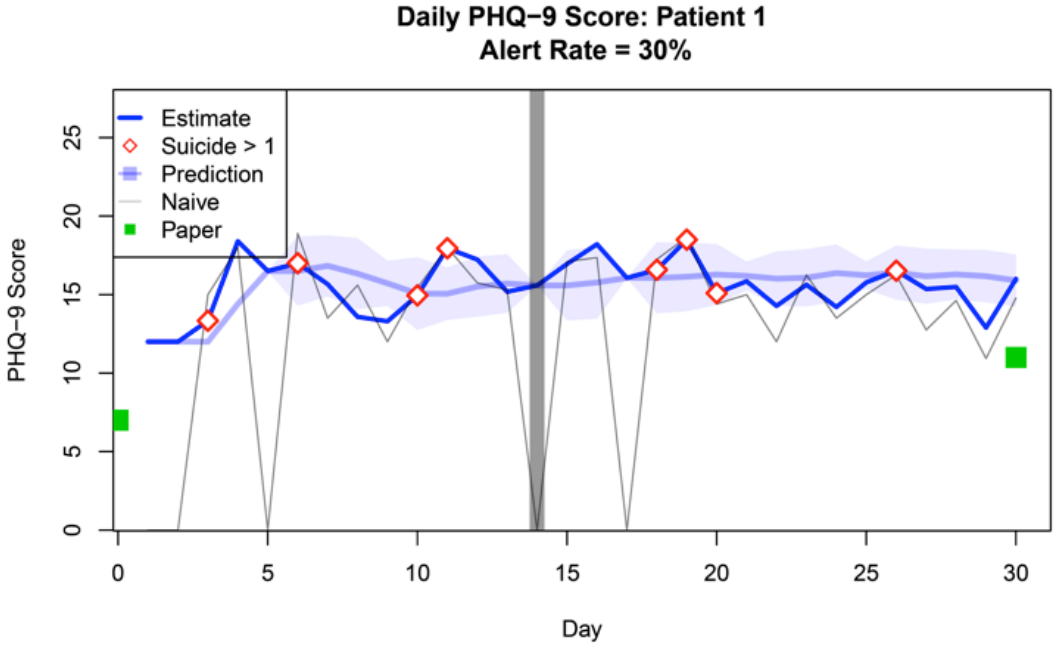
# PILOT STUDY

- **Used Mindful Moods app by John Torous**
- Micro-surveys only
- 13 subjects run for 30 days (29 days for 2 subjects)
- Outpatients with a diagnosis of depression
- 3 micro-surveys per day, 3 questions each, with replacement
- Own phones
- Incentive ($50 for 30 days)


"Utilizing a personal smartphone custom app to assess the Patient Health Questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder" by J Torous, P Staples, M Shanahan, C Lin, P Peck, M Keshavan, and JP Onnela. JMIR Mental Health 2(1), e8 (2015).

# PILOT

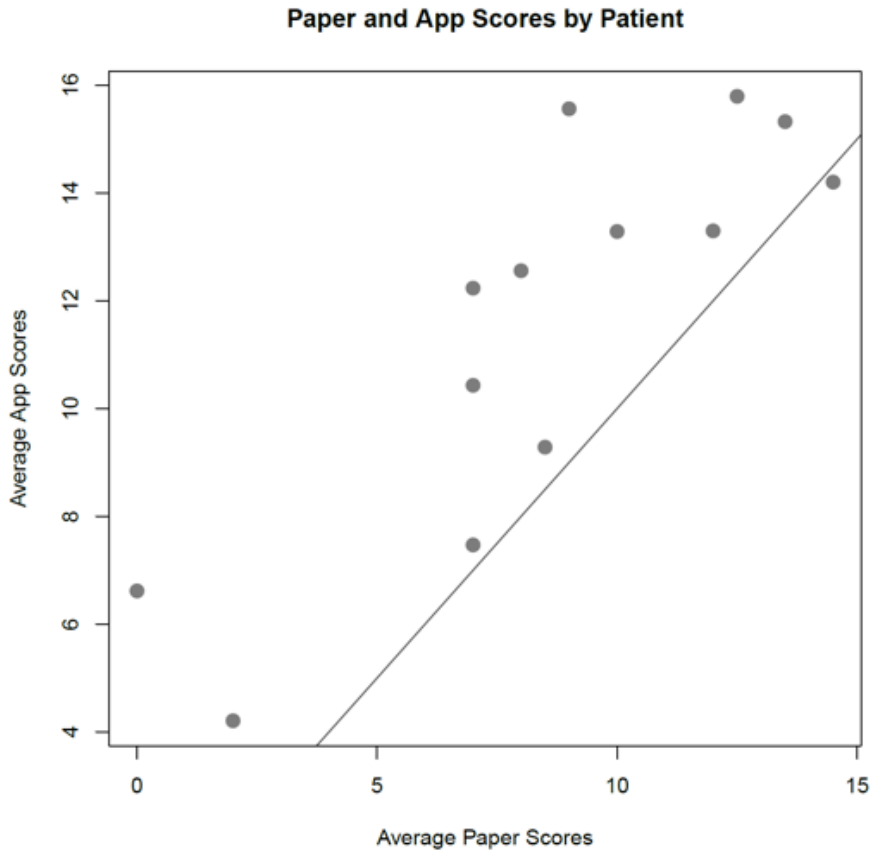



"Utilizing a personal smartphone custom app to assess the Patient Health Questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder" by J Torous, P Staples, M Shanahan, C Lin, P Peck, M Keshavan, and JP Onnela. JMIR Mental Health 2(1), e8 (2015).

# PILOT: PHQ-9 ESTIMATES

- Average app score is on average 3.0 points higher than paper based score
- Average paper and application PHQ-9 scores strongly correlated: Pearson correlation coefficient of 0.84 with 95% CI: (0.55, 0.95)




"Utilizing a personal smartphone custom app to assess the Patient Health Questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder" by J Torous, P Staples, M Shanahan, C Lin, P Peck, M Keshavan, and JP Onnela. JMIR Mental Health 2(1), e8 (2015).

# GPS

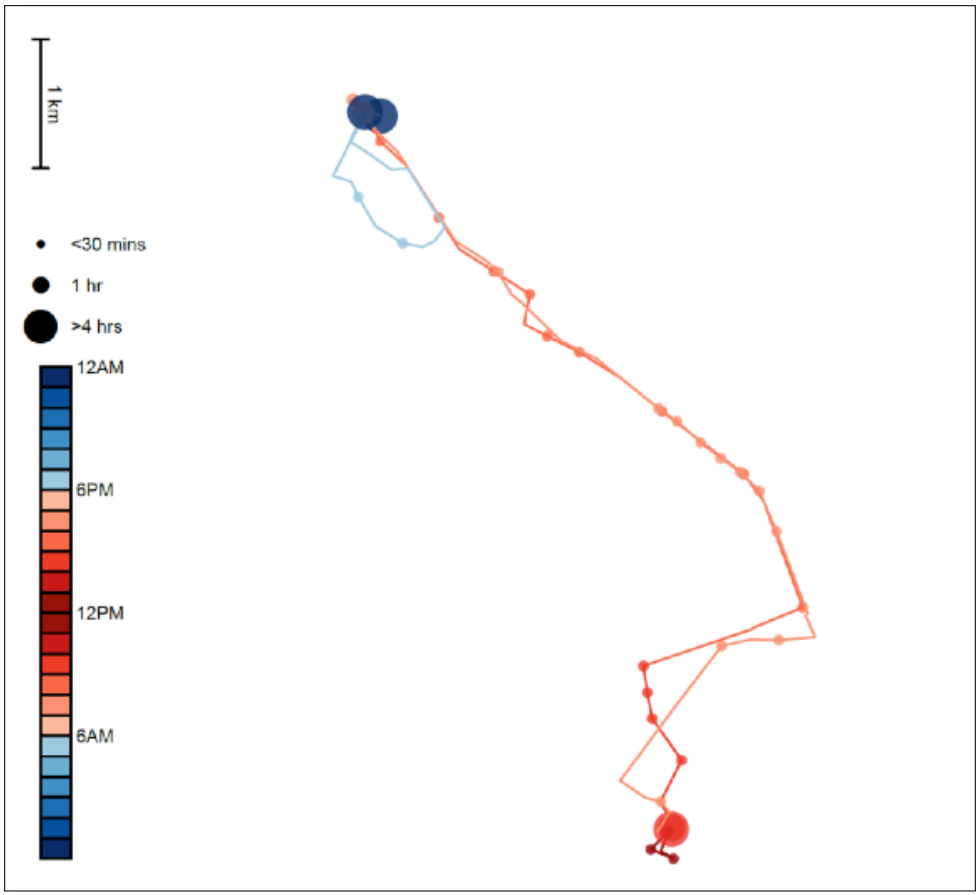
- **Existing approach 1: Use dedicated GPS receivers**
  - Data collected continuously and therefore no missingness
  - Poor scalability and poor long-term adherence

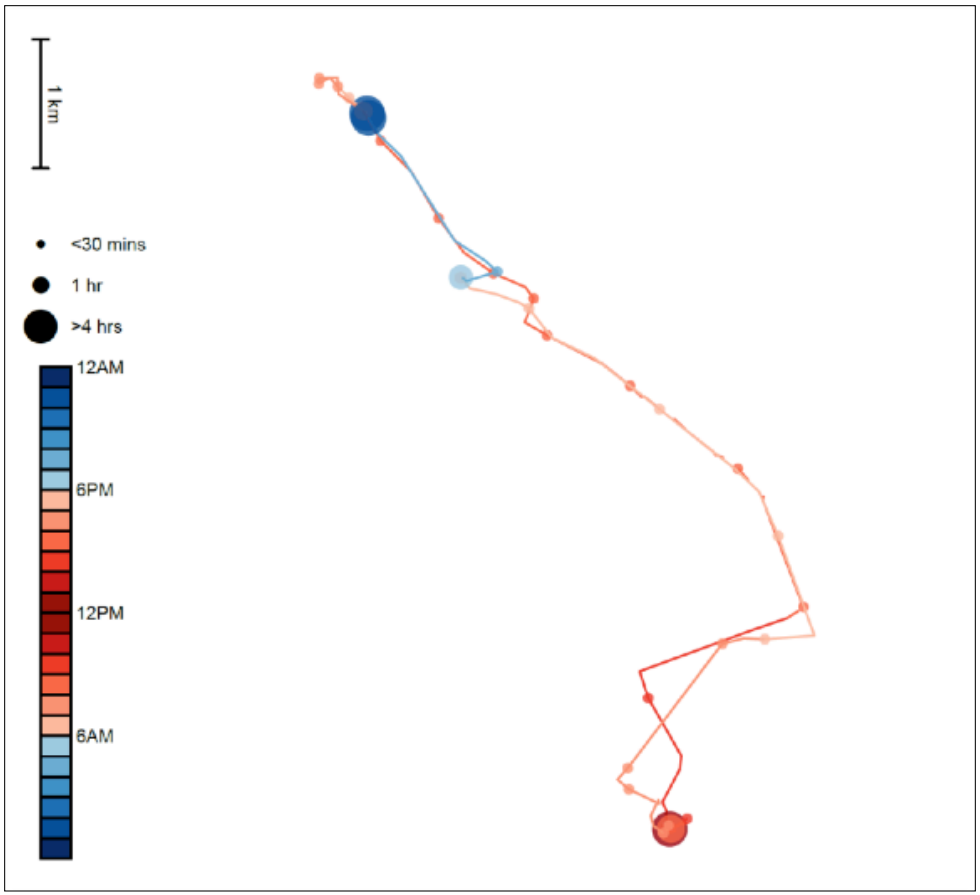
- **Existing approach 2: Use smartphone GPS**
  - Needs to be sampled to avoid battery drainage
  - Ignore missingness or use linear interpolation

- **Our approach: Use smartphone GPS and deal with missingness**
  - Scientific opportunity: scalable for medical and public health applications
  - Statistical opportunity: principled way to deal with missingness

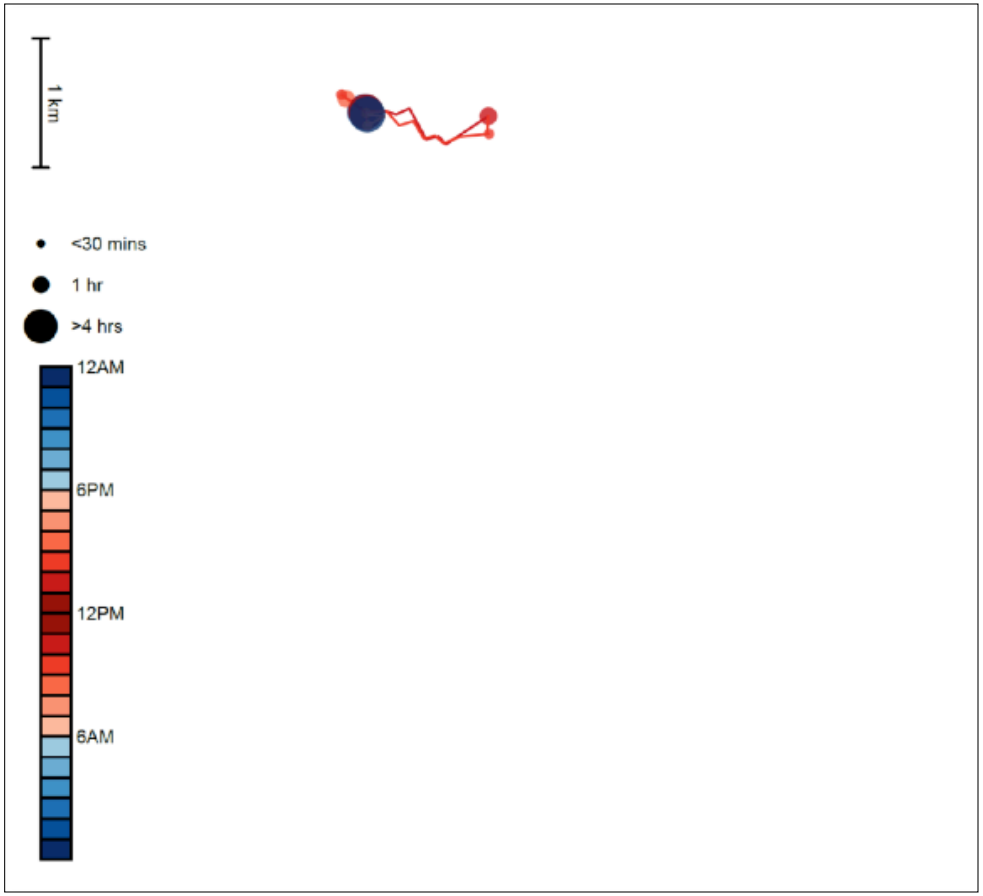
# MOBILITY TRACES

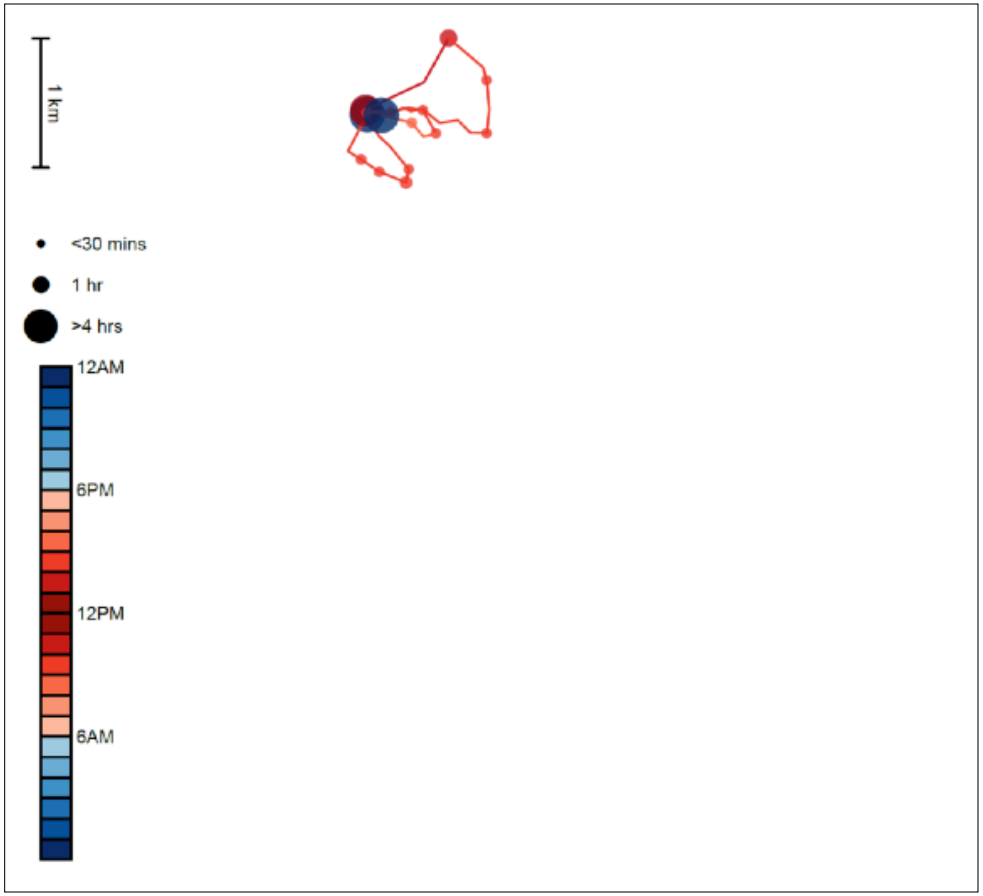
- High frequency GPS trajectory converted to a mobility trace: (1) flights, (2) pauses, (3) time, (4) spatial scale



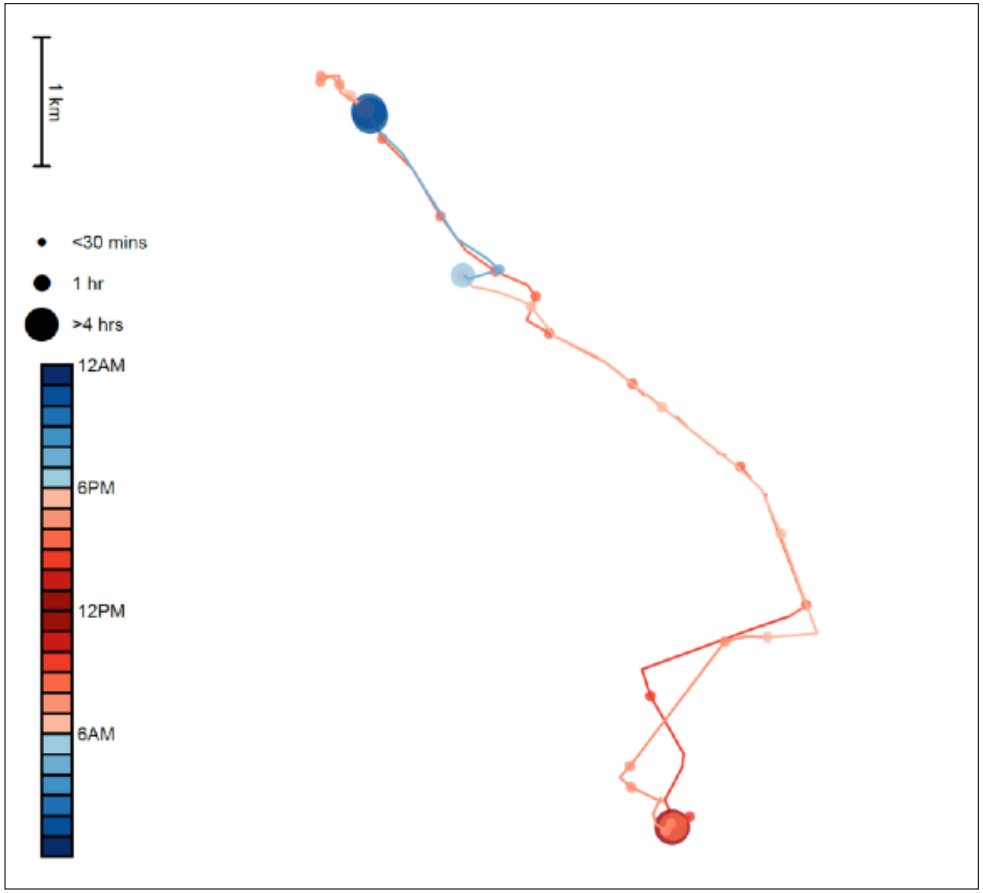
# MOBILITY TRACES

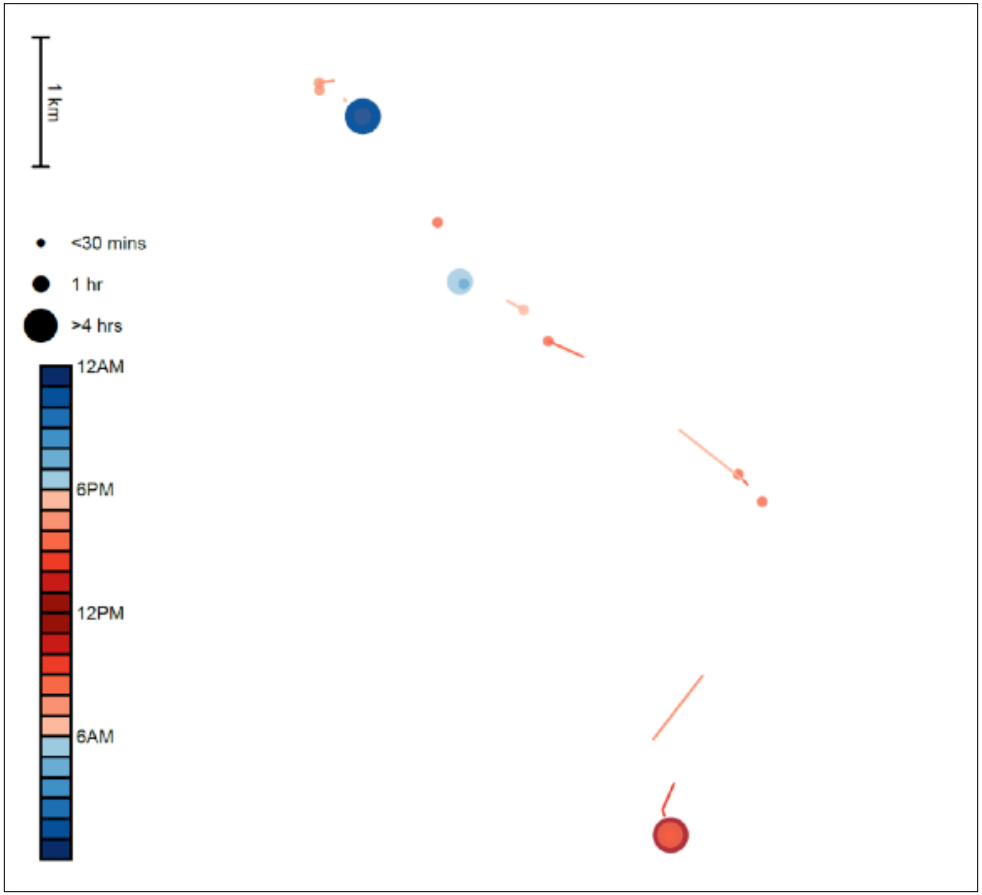


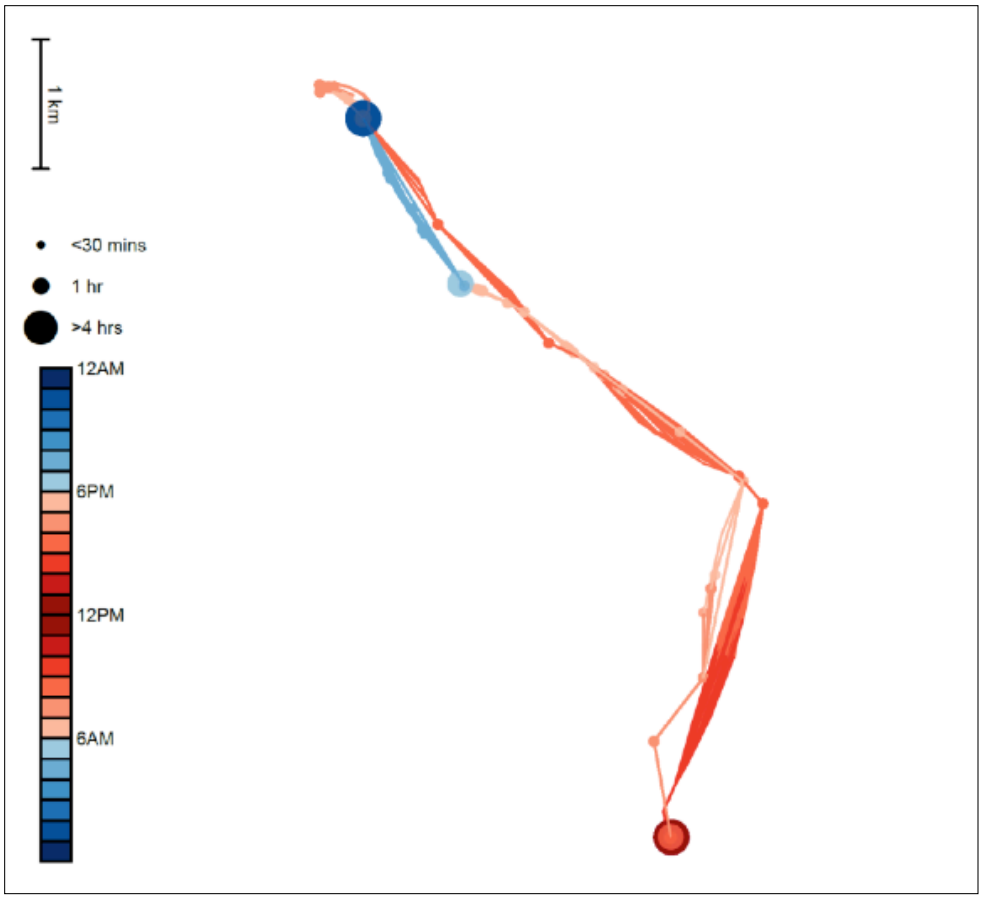
# MISSING DATA

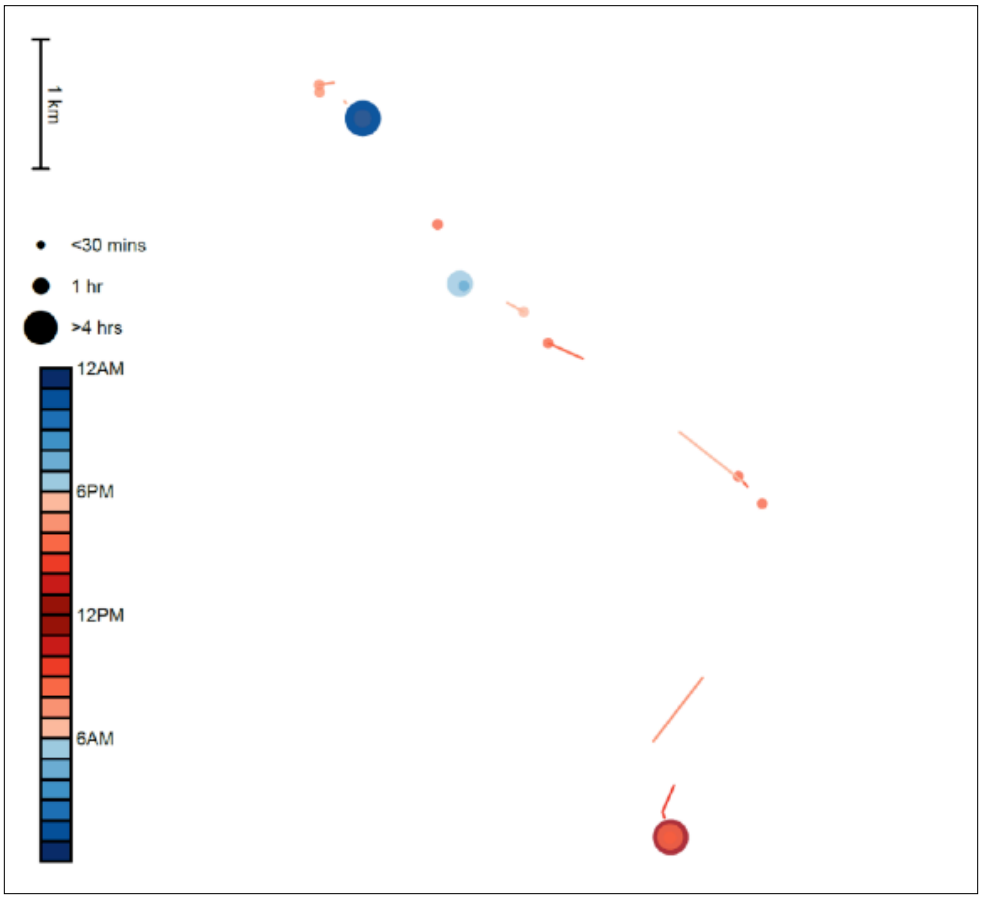
- Complete mobility trace vs. simulated missingness
- Typical sampling cycle on our platform: on-cycle = 2 mins, off-cycle = 10 mins; 83.3% of mobility trace missing



# SIMULATED TRACES



# MOBILITY METRICS

| Measures | TL.1 | TL.10 | TL.20 | GL.1 | GL.10 | GL.20 | GLC.1 | GLC.10 | GLC.20 | LI | Truth |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Hometime | 831.5 ±2.3 | 832.3 ±2.4 | 833.4 ±2.2 | 830.3 ±2.2 | 830.5 ±2.8 | 829.8 ±1.9 | 829.1 ±2.1 | 832.1 ±2.2 | 831.3 ±2.5 | 826.7 | 882.8 |
| DistTravelled | 22184 ±969.7 | 22446 ±843.5 | 22569 ±811.6 | 18801 ±466.3 | 18801 ±337.5 | 18779 ±369.4 | 21791 ±969.9 | 22380 ±712.1 | 22444 ±645.6 | 17236 | 19344 |
| RoG | 2787.3 ±2.3 | 2791.3 ±2.6 | 2791.2 ±1.9 | 2783.0 ±1.6 | 2783.0 ±1.9 | 2783.3 ±2.5 | 2785.6 ±1.3 | 2787.0 ±1.5 | 2787.5 ±1.8 | 2779.4 | 2781.3 |
| MaxDiam | 6717 ±169 | 6745 ±129 | 6727 ±98 | 6494 ±44 | 6483 ±8 | 6496 ±34 | 6516 ±55 | 6517 ±55 | 6562 ±94 | 6479 | 6467 |
| MaxHomeDist | 6372 ±165 | 6410 ±123 | 6379 ±93 | 6160 ±49 | 6147 ±16 | 6153 ±39 | 6144 ±30 | 6152 ±5 | 6163 ±24 | 6149 | 6129 |
| SigLocsVisited | 2.96 ±0.73 | 3.20 ±0.58 | 3.20 ±0.71 | 3.16 ±0.69 | 3.00 ±0.76 | 2.96 ±0.79 | 3.28 ±0.61 | 3.12 ±0.60 | 3.20 ±0.65 | 2 | 3 |
| AvgFlightLen | 172.7 ±10.7 | 160.2 ±7.6 | 158.6 ±7.4 | 200.2 ±23.2 | 193.2 ±19.2 | 191.7 ±18.1 | 129.9 ±13.6 | 122.8 ±6.1 | 127.1 ±7.6 | 478.8 | 251.2 |
| StdFlightLen | 152.9 ±30.8 | 125.8 ±10.1 | 123.2 ±5.5 | 213.4 ±51.5 | 205.8 ±36.3 | 202.7 ±43.5 | 151.0 ±30.0 | 134.2 ±8.4 | 137.1 ±9.0 | 639.6 | 223.3 |
| AvgFlightDur | 79.0 ±9.3 | 69.4 ±5.8 | 68.8 ±5.6 | 119.0 ±17.9 | 115.2 ±13.4 | 113.5 ±13.7 | 65.4 ±10.5 | 57.2 ±4.1 | 60.0 ±5.1 | 340.6 | 77.0 |
| StdFlightDur | 131.7 ±17.0 | 115.3 ±9.0 | 113.5 ±10.2 | 170.3 ±22.0 | 168.7 ±14.8 | 166.7 ±14.4 | 103.7 ±18.2 | 85.0 ±10.9 | 91.7 ±13.1 | 289.8 | 55.2 |
| FracPause | 0.88 ±0.01 | 0.89 ±0.01 | 0.89 ±0.01 | 0.87 ±0.01 | 0.87 ±0.01 | 0.87 ±0.01 | 0.87 ±0.01 | 0.88 ±0.01 | 0.88 ±0.01 | 0.86 | 0.93 |
| SigLocEntropy | 0.63 ±0.01 | 0.63 ±0.01 | 0.63 ±0.01 | 0.63 ±0.01 | 0.63 ±0.01 | 0.63 ±0.01 | 0.63 ±0.01 | 0.63 ±0.01 | 0.63 ±0.01 | 0.63 | 0.63 |
| MinsMissing | 1243 | 1243 | 1243 | 1243 | 1243 | 1243 | 1243 | 1243 | 1243 | 1243 | 92 |
| CircdnRtn | 0.64 ±0.02 | 0.63 ±0.01 | 0.63 ±0.02 | 0.67 ±0.01 | 0.67 ±0.01 | 0.67 ±0.01 | 0.65 ±0.02 | 0.66 ±0.01 | 0.66 ±0.02 | 0.69 | 0.66 |
| WkEndDayRtn | 0.76 ±0.02 | 0.76 ±0.01 | 0.76 ±0.01 | 0.78 ±0.01 | 0.77 ±0.01 | 0.78 ±0.01 | 0.76 ±0.02 | 0.76 ±0.01 | 0.77 ±0.01 | 0.81 | 0.79 |

# THANK YOU !

**www.hsph.harvard.edu/onnela-lab/**