AN AD-HOC APPROACH TO DATA QUALITY ASSESSMENT: THE BIKING & BREATHING STUDY IN NYC

Darby Jack

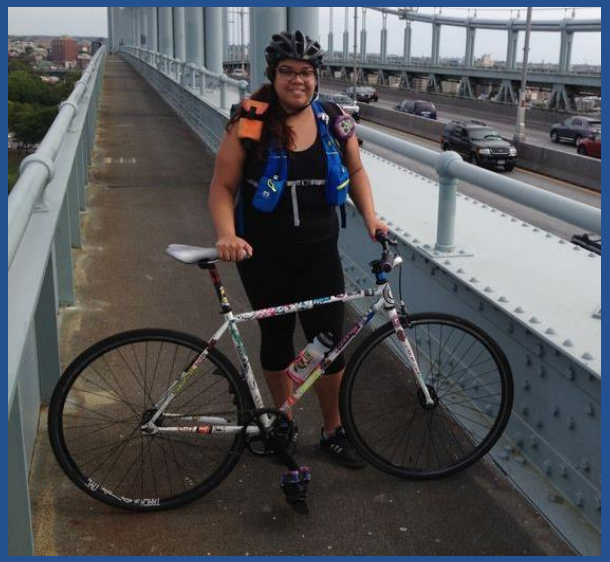
dj2183@cumc.Columbia.edu

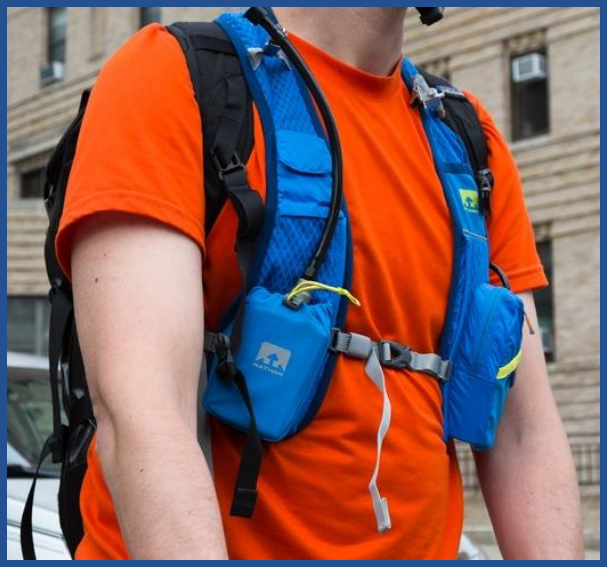
EPA Air Sensors 2018

Biking and Breathing study design

- In partnership with WNYC, we recruit bike commuters who ride at least 30 minutes each way.
- Volunteers self-deploy sensors for six 24 -hour monitoring sessions bracketing a morning commute.
- We estimate potential inhaled dose using minute ventilation from a biometric shirt.
- Epi hypothesis: short duration air pollution exposures increase post-exposure BP and decrease heart rate variability.
- Currently in year 4 of a 5 year study.
- Today:
  - How do we think about data quality in the core epi study?
  - How did we think about data quality in a pilot evaluation of low cost sensors?
  - How might we do a better job?

Environmental Monitors





MicroAeth
Black Carbon

Smart Phone App for GPS

MicroPEM
$PM_{2.5}$

$PM_{2.5}$ data quality – core epi study

The RTI microPEM gives real-time optical $PM_{2.5}$ estimates, and also collects a gravimetric filter.

- Objective 1: sensor was deployed properly & functioned properly
  - Wearing compliance (via accelerometer)
  - Sampling duration
  - Stable flow rate
  - RH <90%
- Objective 2: plausible data
  - Reasonable baseline drift
  - Clean air test at the start and finish of each 24-hour deployment
  - No negative values, implausibly high values
- Objective 3: consistent with other measurements
  - Plausible BC/PM ratio
  - Mass concentration from gravimetric filter (one correction factor for 6 x 24 hours)
  - Field blanks, duplicates

Extensions: pilot study evaluating low cost sensors

- What if low cost, small, self deployed sensors could collect high quality data?
- We pilot-tested a software + hardware system that uses
  - knowledge about sensor physical properties,
  - coincidental colocations (with regulatory monitors, and among low cost sensors themselves)
- Currently ingests data from AirBeam ($PM_{2.5}$; $250) and Terrier ($CO/NO/CO_2$; $330).
- Evaluation approach:
  - Comparison to microPEM monitors – ask volunteers to carry both for approx. 30 days.
  - Mass deployments (group rides with 10+ monitors)

Low cost sensor evaluation

- Primary data quality objective: correlation w/ microPEM.
  - Overall the low cost monitor did not perform well and postprocessing system did not result in material improvements
  - RMSE = 13 $\mu g/m^3$ (grand mean: ~10 $\mu g/m^3$); R^2 = 27%.
  - Minute-by-minute error was not significantly correlated with temperature or RH.
  - Caveat: microPEM not a "gold standard"
- Secondary data quality objective: correlation between low cost monitors in group rides. Similarly low.
- We did not have a pre-defined R^2 threshold for declaring success, but it was clear that the low cost monitor was not reliable enough for use in epi studies.

Where to from here?

- Data quality assessment is not a one-off task
  - **Assess a deployment, not a device**
  - Rationale: inter-deployment changes in the sensor, changes in meteorology (temp, RH), changes in particle composition, size distribution.
- Stratified data quality assessments:
  - Meteorological conditions (temp, RH)
  - PM concentration quantiles
- Optical measurements need external checks
  - Continued role for gravimetric sampling (correction factors)
  - Simple but effective: colocation, clean air check (zero box).
- Use automated tools for visualizing data & generating statistics

  (if it's inconvenient it doesn't happen) → data standards
- **Predefine** clear rules of thumb for accepting/rejecting data (quantitative but not necessarily optimal – fit for purpose).

Acknowledgements

Co PI: Steve Chillrud

Funding: NIEHS (R21/R33 ES 024734, P30 ES 009089)

Collaborators: Pat Kinney, Aimee Layton, Ali Lee, Daichi Shimbo, Richard Sloan, Cara Smith, Jon Thornburg, Qiang Yang.

WNYC (Fred Mogul)

www.wnyc.org/streets

Our amazing study participants who tolerate way too many sensors.