Boston University School of Public Health

Bearing Witness: Using Digital Data to Monitor Health & Disease Patterns

Elaine Nsoesie, PhD

Assistant Professor of Global Health

Email: onelaine@bu.edu / Twitter: @ensoesie

Digital Epidemiology

Epidemiology is the study of the patterns of disease and health in populations and the factors that contribute to these patterns.

Digital epidemiology is the same but with digital data

Types of data & uses

Crowdsourcing

Search

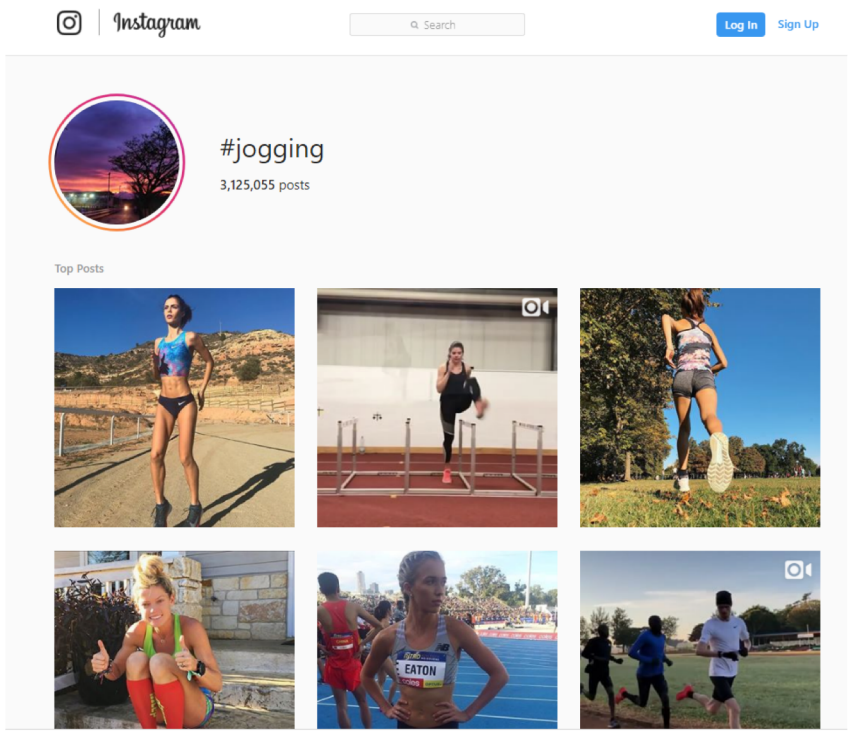
Social media

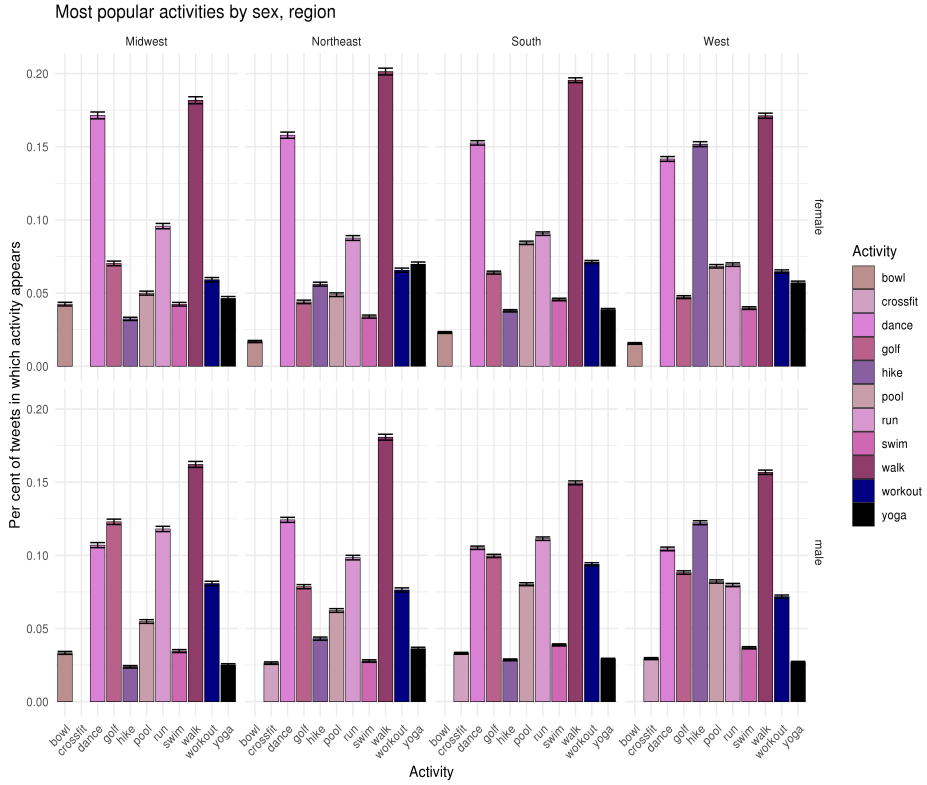
Consumer reviews

Remote sensing/place

News

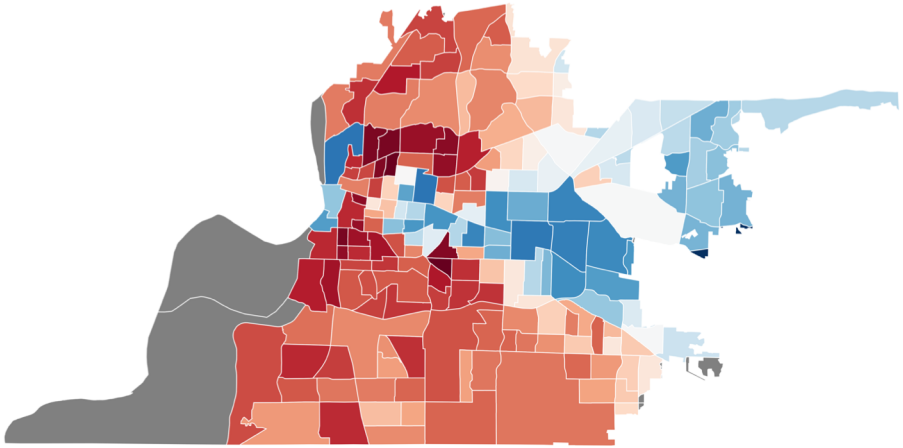
Example 1: Monitoring Physical Activity





Example 2: The Built Environment and Obesity

Memphis, Tennessee



Limitations and Future Research Directions

Ethics and privacy

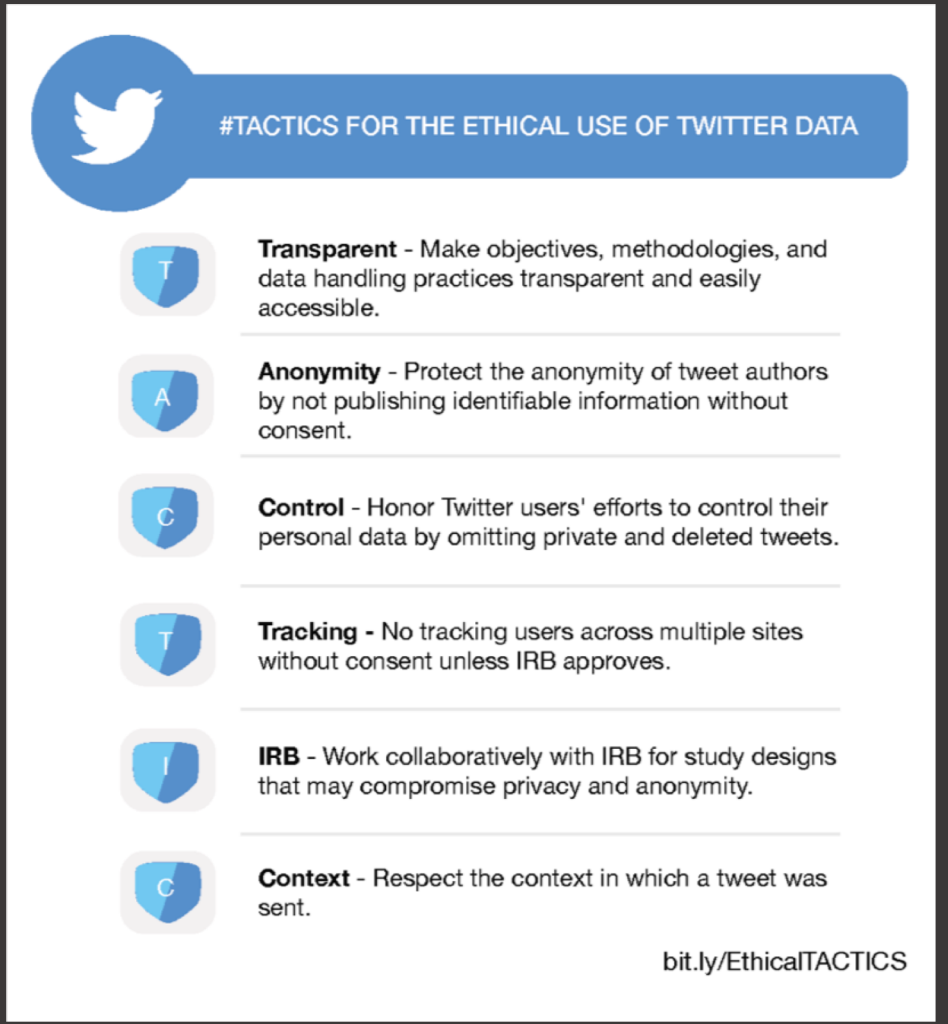
Human subjects standards and ethical norms are not well-established

The Institutional Review Board

IRBs aim to protect research subjects from the potential harms caused by the research methods.

"Pervasive datasets dramatically change the research ethics landscape. The research methods and risks have changed, but the regulations have not." – Jacob Metcalf

IRB approval does not imply your study is ethical



Demographic and Spatial Representation

Data is sometimes biased; not everyone is represented

Name: Angelina
Age: 45
Number of jobs: 3
Health Insurance: None
Social Media Use: Facebook (on occasion)

Photo by Gabrielle Henderson on Unsplash

Photo by Samantha Gades on Unsplash

3 Data Sources

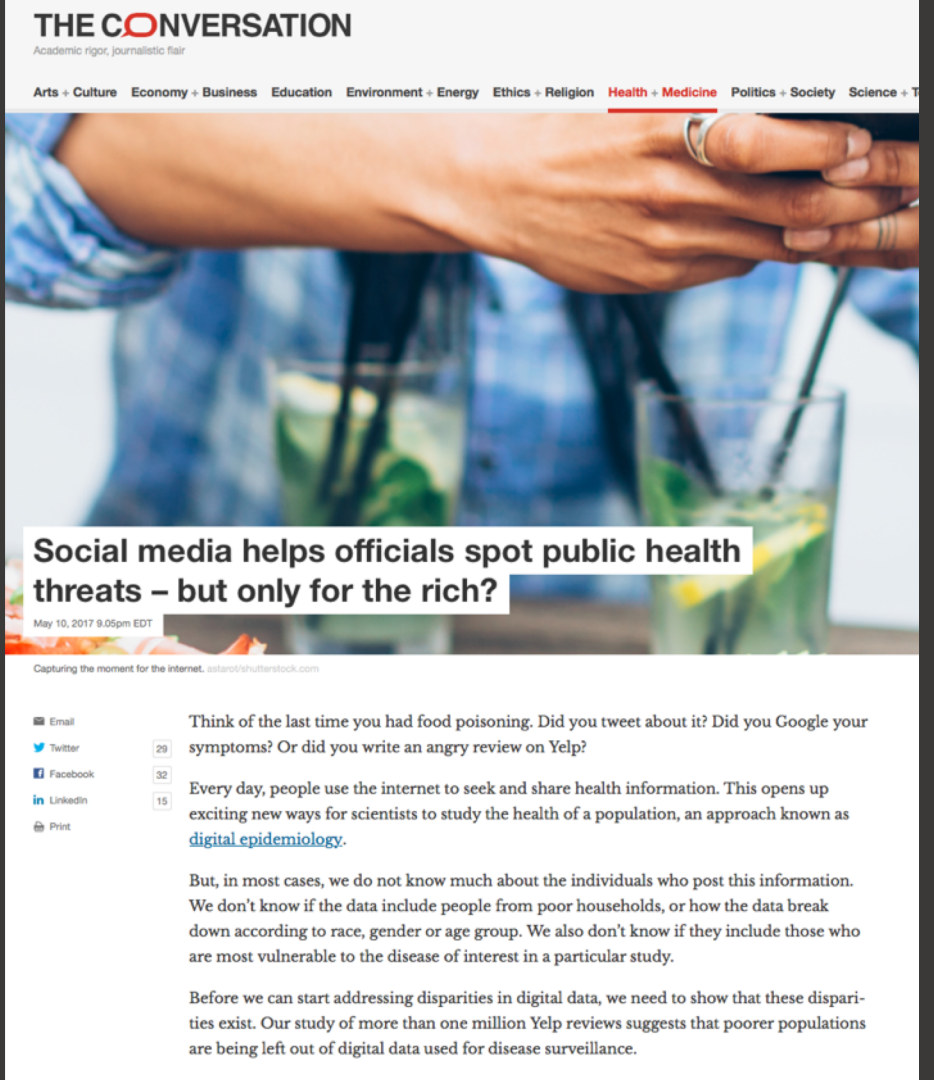
- Traditional ILI surveillance system (ILINet)
- BioSense 2.0
- Google Flu Trends

“The models make the best predictions in the most affluent zip codes and the worst predictions in the most impoverished zip codes, regardless of the data sources."

Scarpino et al. (2018) *Socioeconomic bias in influenza surveillance. arXiv*

yelp

Health departments in the US are using data from social media for foodborne illness surveillance



THE CONVERSATION

Academic rigor, journalistic flair

Arts + Culture | Economy + Business | Education | Environment + Energy | Ethics + Religion | Health + Medicine | Politics + Society | Science + T

Social media helps officials spot public health threats – but only for the rich?

May 10, 2017 9.05pm EDT

Capturing the moment for the internet. astarot/shutterstock.com

Email
Twitter 29
Facebook 32
LinkedIn 15
Print

Think of the last time you had food poisoning. Did you tweet about it? Did you Google your symptoms? Or did you write an angry review on Yelp?

Every day, people use the internet to seek and share health information. This opens up exciting new ways for scientists to study the health of a population, an approach known as digital epidemiology.

But, in most cases, we do not know much about the individuals who post this information. We don't know if the data include people from poor households, or how the data break down according to race, gender or age group. We also don't know if they include those who are most vulnerable to the disease of interest in a particular study.

Before we can start addressing disparities in digital data, we need to show that these disparities exist. Our study of more than one million Yelp reviews suggests that poorer populations are being left out of digital data used for disease surveillance.

Yelp Reviews

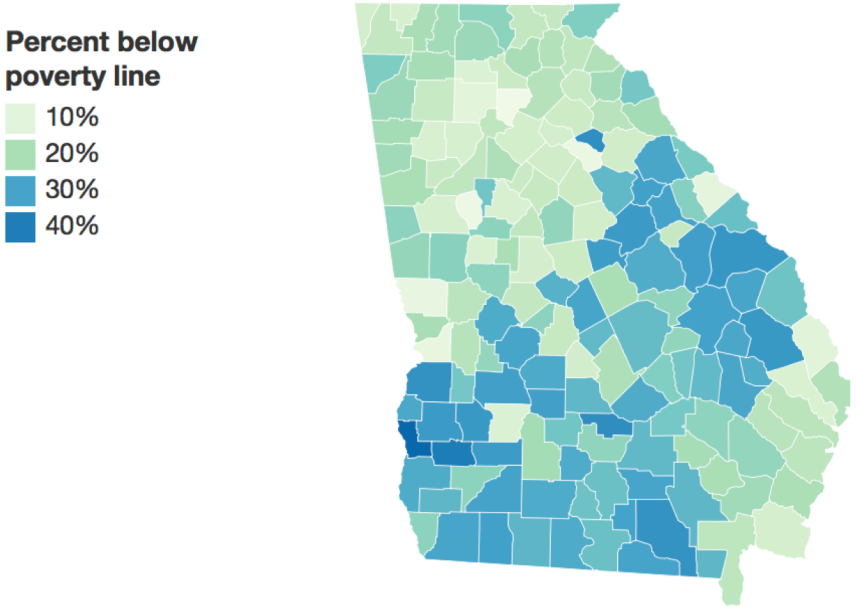
- 2004-2014
- 1.5 million foodservice business reviews
- Oregon, Georgia, and Massachusetts

“Factors typically associated with affluence (such as higher income and having a bachelor’s degree) were positively correlated with foodborne illness reports."

Henly et al. *Preventive Medicine*, (2017)

Using Yelp for foodborne illness surveillance

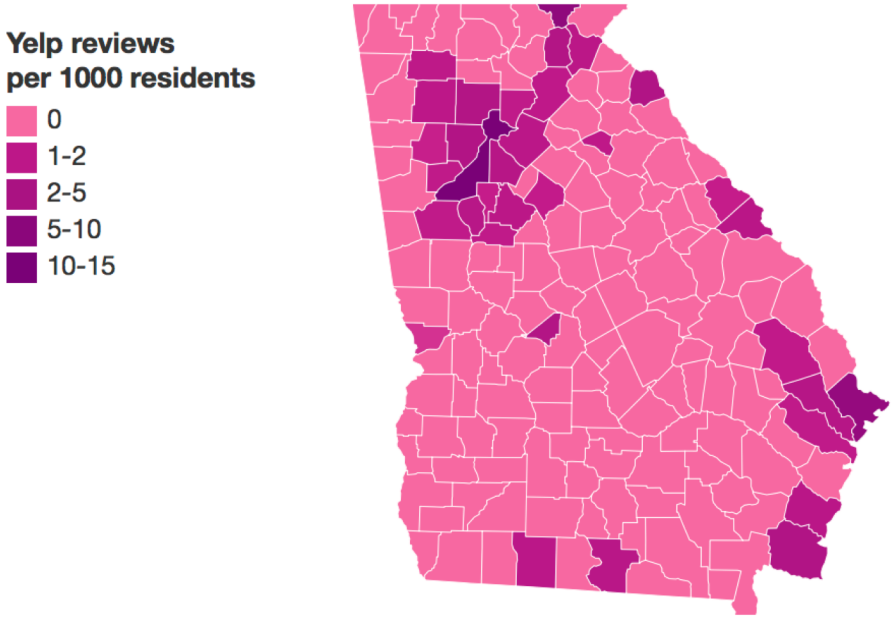
Poverty in Georgia



The Conversation, CC-BY-ND

Source: Elaine Nsoesie Get the data

Yelp reviews in Georgia



The Conversation, CC-BY-ND

Source: Elaine Nsoesie Get the data

Methods

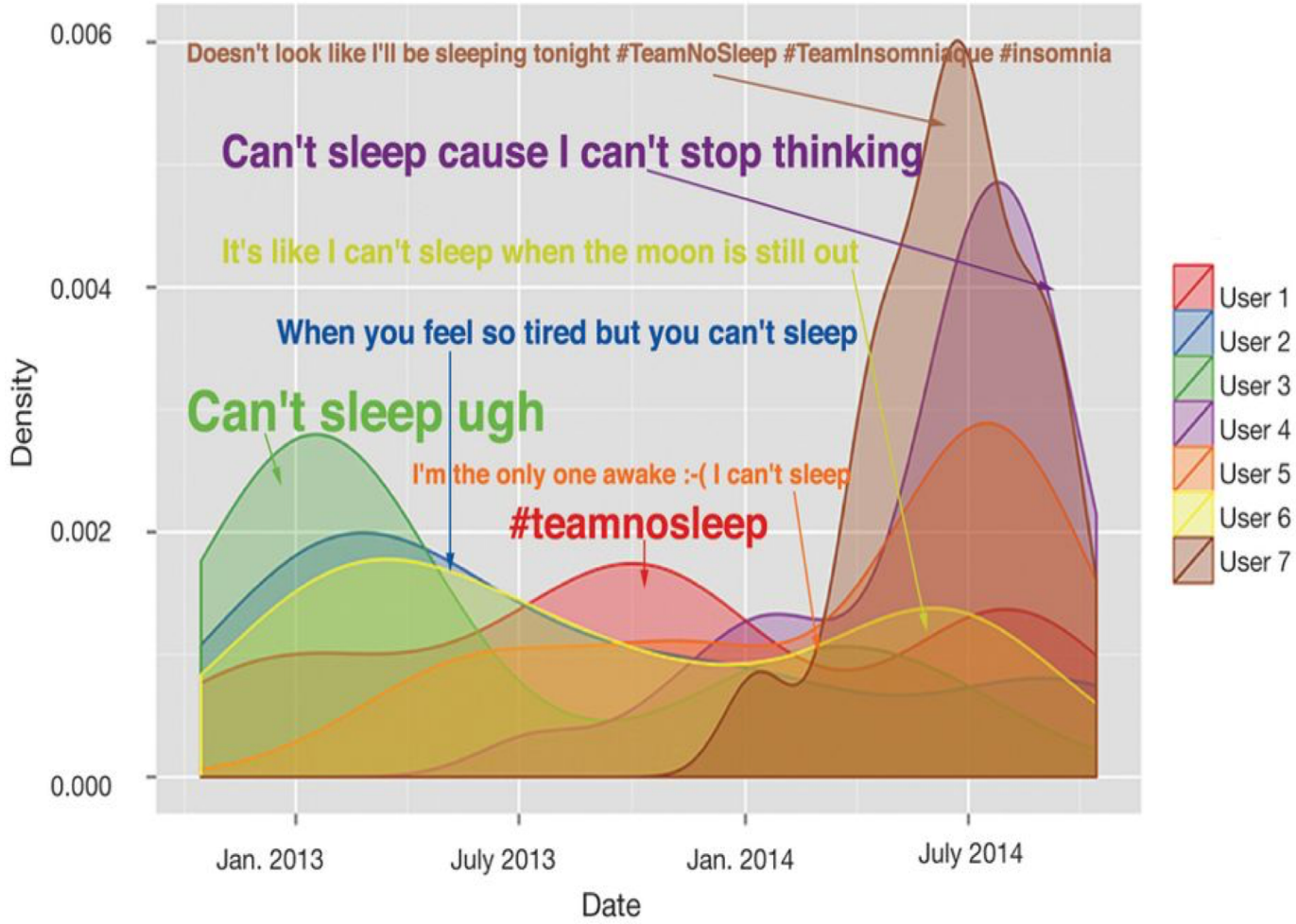
Analytical methods have not been tested, validated and standardized across research domains

We need **methods for rigorous analysis** of various streams of digital health data

These would enable us to identify what is useful and what isn't

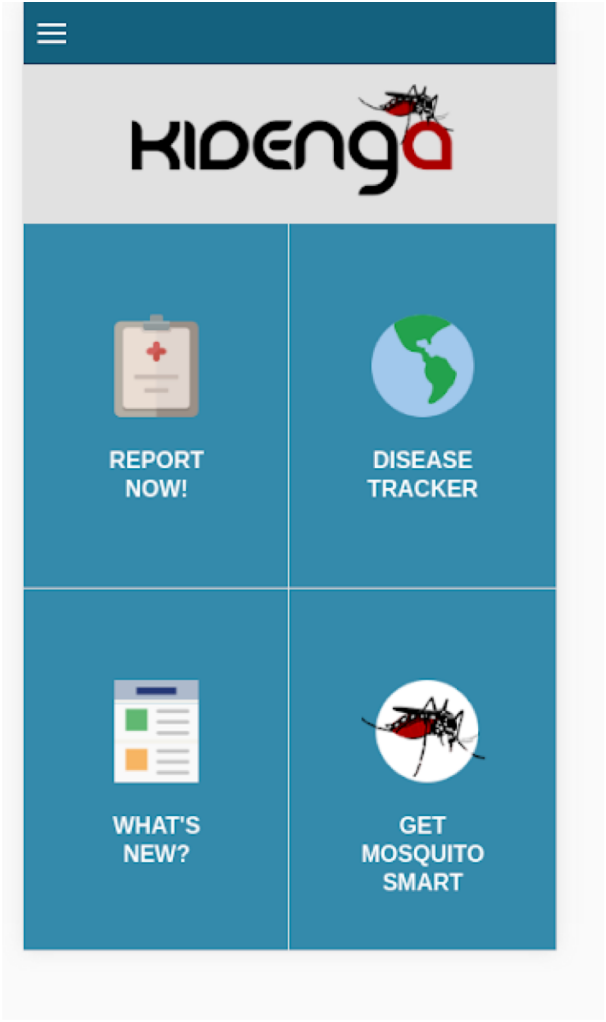
And also to quantify uncertainty

Jain et al. *Nature Biotechnology*, (2015)



We need new tools that would allow for identification of useful approaches and **easy integration of different data streams**

We need to develop methods and informatics tools that process, **integrate and make available these data to local decision makers for public health surveillance**



Robert Wood Johnson Foundation

Contact
Email: onelaine@bu.edu
Twitter: @ensoesie