

# **The Parable of Google Flu: Traps in Big Data Analysis**

David Lazer, Ryan Kennedy, Gary King, Alessandro Vespignani

Arnold Cieremans

# What it is about

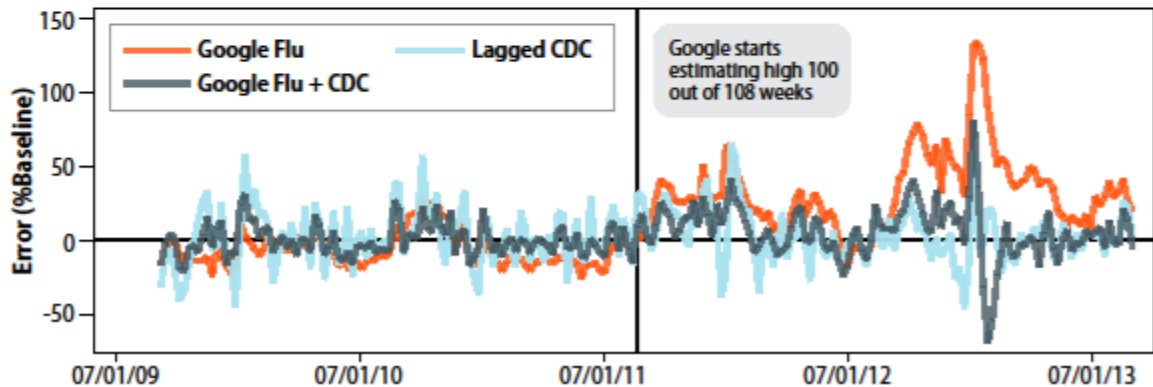
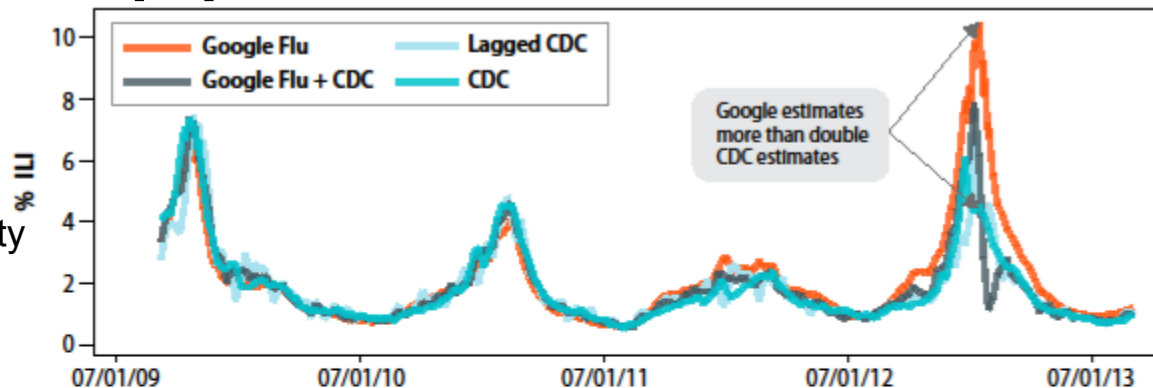
- Google Flu Trends (GFT)
- Centers for Disease Control and Prevention (CDC)
- Serves as example
- What is broken?
  - Big data hubris
  - Algorithm dynamics
- How to fix?

# Big data hubris

- Seen as a substitute
- Unsuitable
- 50 million search terms
- 2009 update

# Big data hubris (2)

- 100 out of 108 weeks
- Temporal autocorrection, seasonality
- Lagged models
- Rendered useless?



# Algorithm dynamics

- Media-stoked panic
  - Bird flu/Swine flu
- Not a static entity
- Google Correlate
- Blue team dynamics
  - Business model

# Algorithm dynamics (2)

- Recommended searches
- External events
- Twitter, Facebook
- Red team dynamics

# What do they recommend

- Transparency, replicability
  - The data is there
- Understand the unknown
  - Granular view
- Study the algorithm
  - Better understanding
- Not just about the size of data
  - Merge small and big
- All data revolution
  - Innovative analytics