Big Data Science: Does it Apply to Public Health?

Dean Ayman El-Mohandes, MBBCh, MD, MPH
CUNY Graduate School of Public Health and Health Policy

CUNY SPH GRADUATE SCHOOL OF PUBLIC HEALTH & HEALTH POLICY

Big Data

The definition does not refer to the amount of data but to data that cannot be stored and processed using traditional data-based software, e.g., Google search index and Facebook user profile.

Big Data

Big data floods businesses on a daily basis, which incentivized them to analyze this data for insights helpful in strategic business decisions.

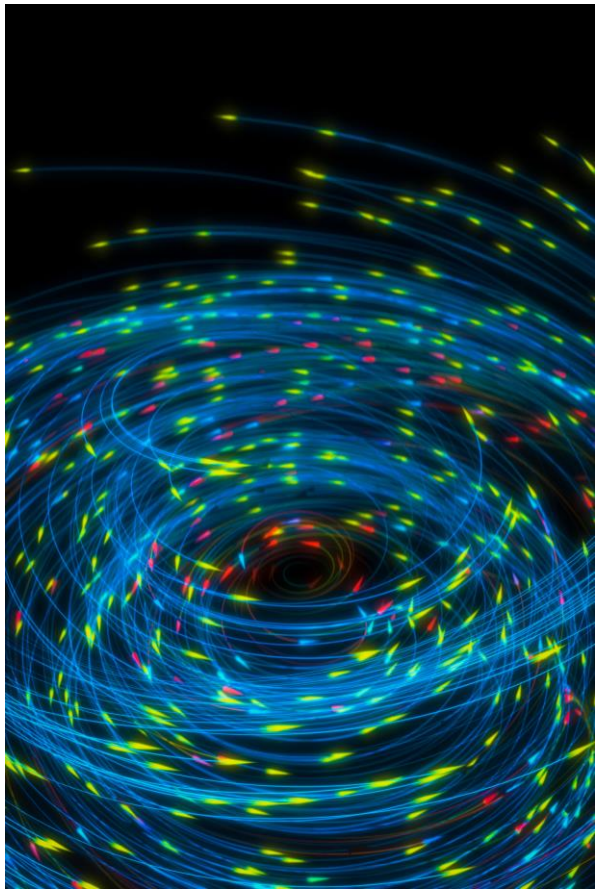
Big Data

Researchers can link disparate data sets to glean insights and interpret information that is predictive, exploratory, or suggestive of complex risk associations and unanticipated outcomes.

Big Data

The Wall Street Journal's initial prediction that big data research would be the demise of traditional statistical research has not panned out to be true.

Characteristics of big data:



- Volume
- Variety
- Velocity

CUNY SPH

Newly-added characteristics:

IBM — Veracity

sas — Variability and Complexity

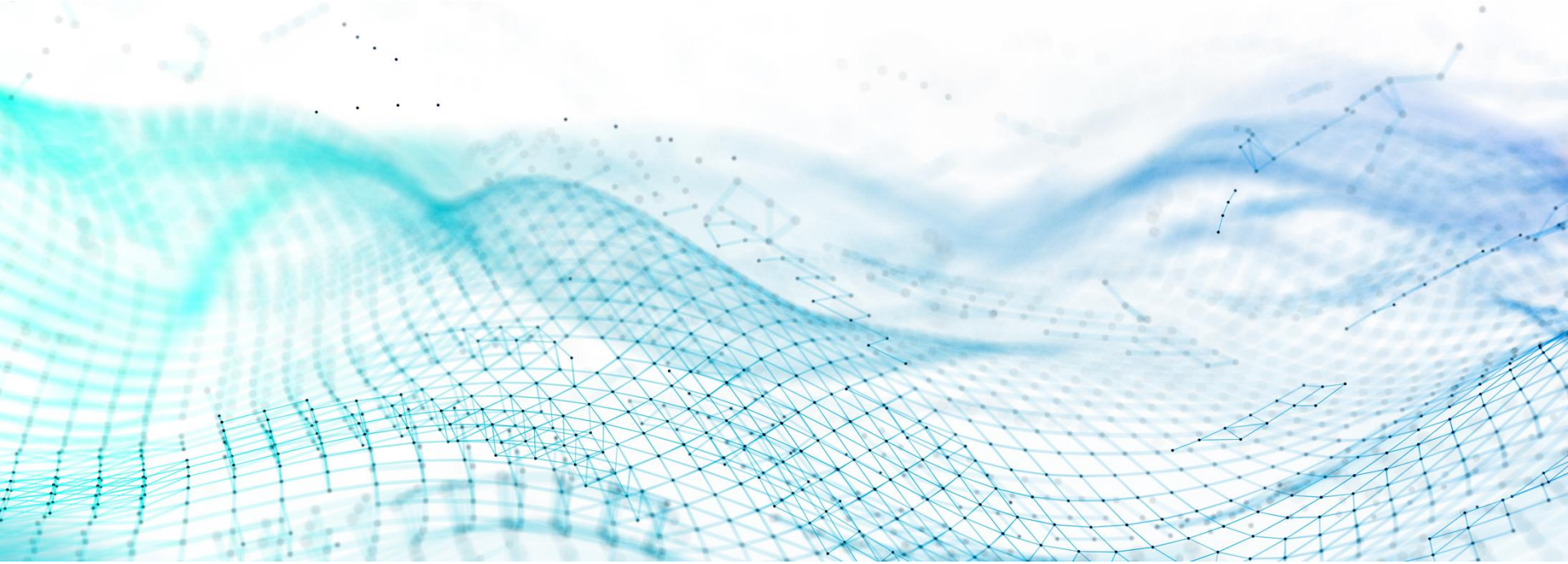
ORACLE — Value

Processing big data:

- Acquiring and Recording
- Extraction and Cleaning
- Integration and Aggregation
- Modeling and Analysis
- Interpretation

Challenges:

Linking disparate data sets is associated with complex technical and ethical challenges.



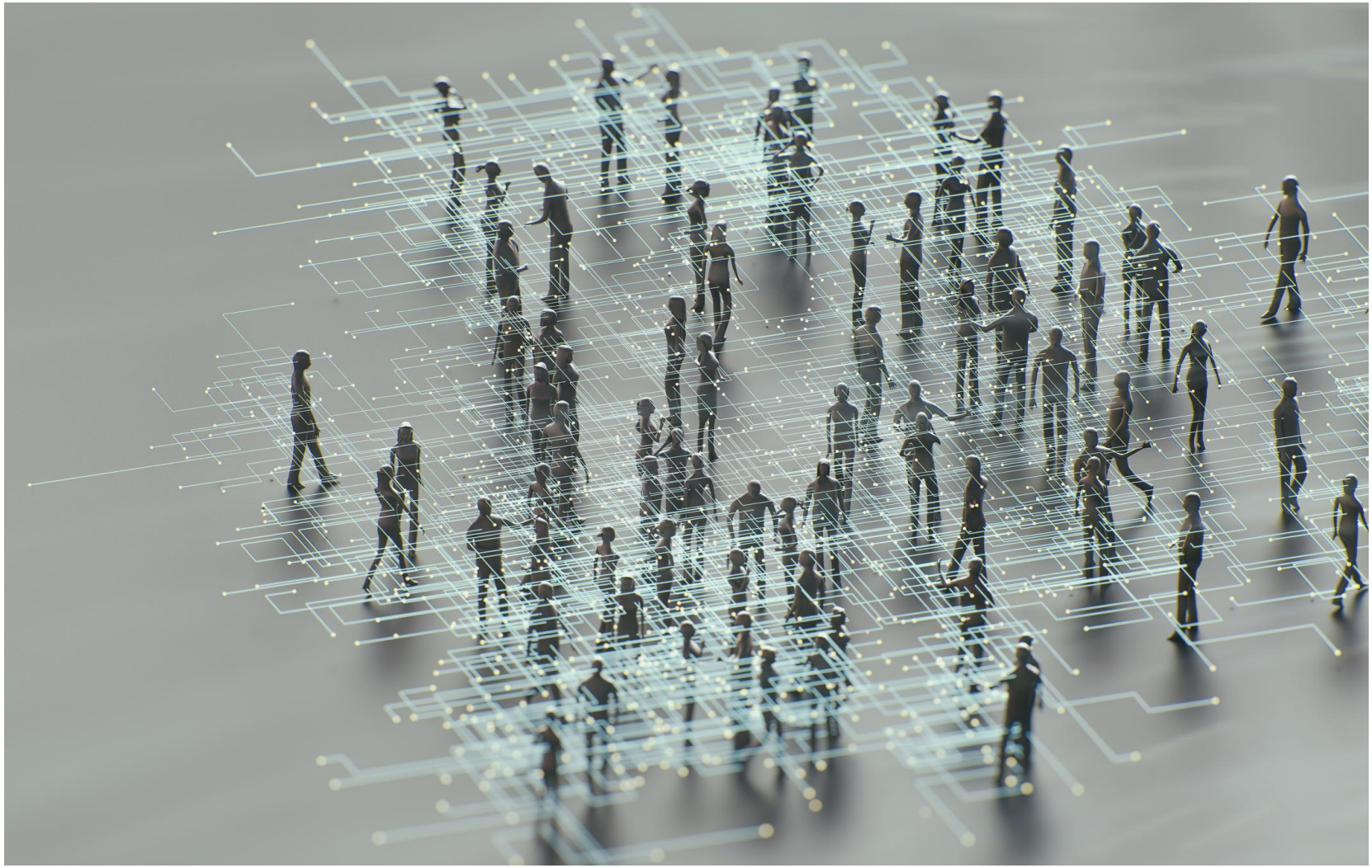
Challenges:

- Loss of representativeness
- Lack of common identifiers
- Issues related to respondent consent and data privacy

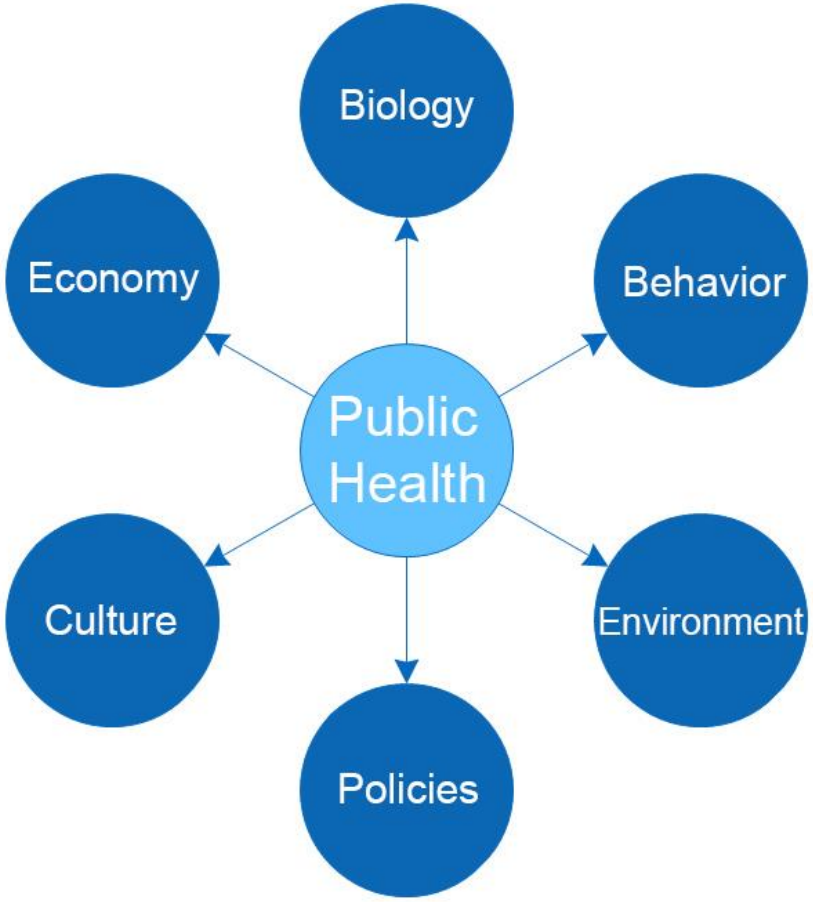
Response:

- New matching technologies to increase the likelihood of correct linkages
- Methods that balance the cost of false positives and false negatives

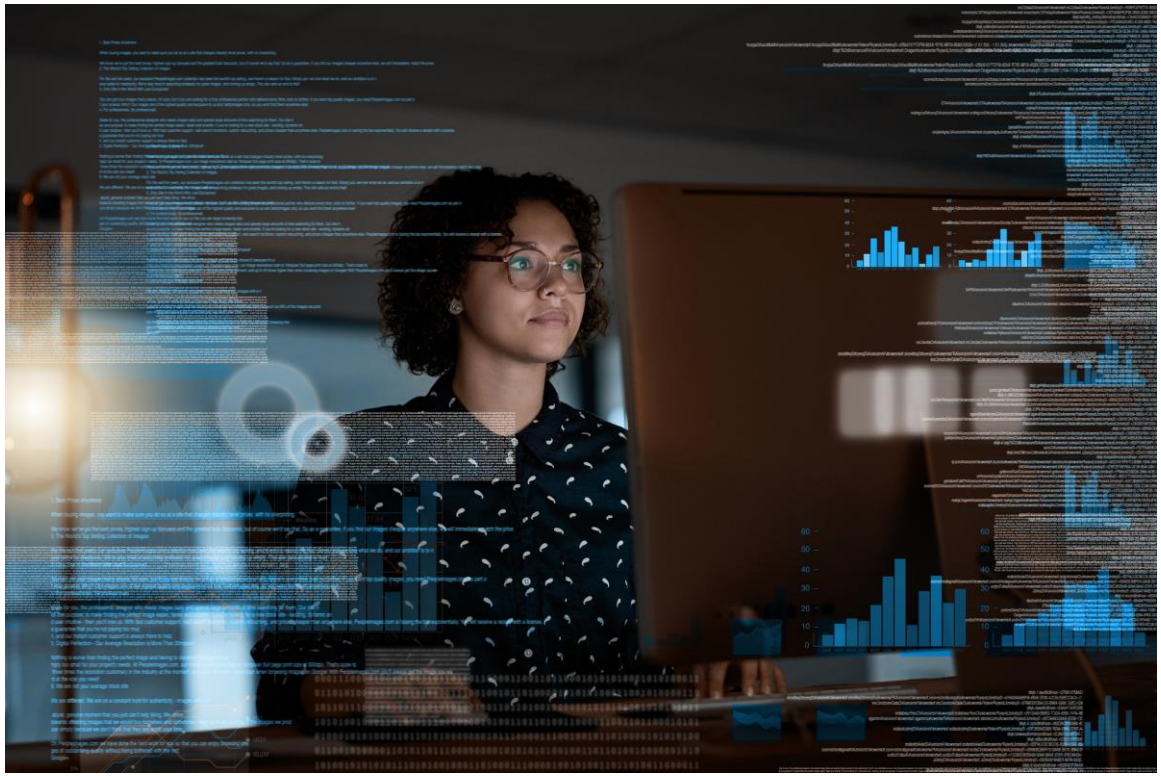
Why is big data research applicable to public health?



Major public health challenges respond to a variety of individual and ecological factors:

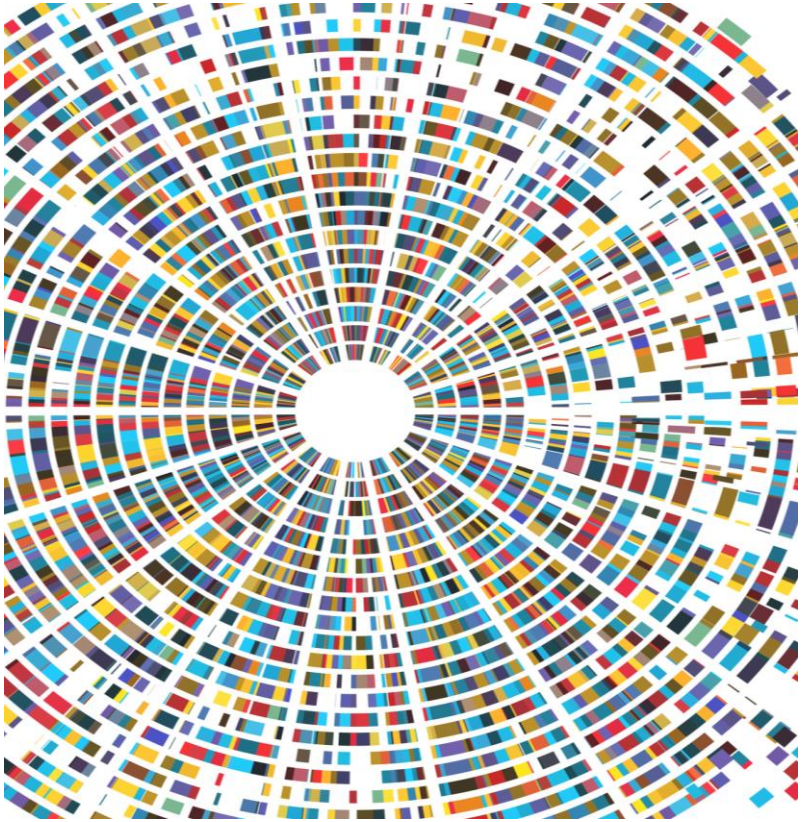


Using a systems application reliant on big data analysis helps save time in discovering unintended consequences or associations and avoidance of linear interpretation of causality.



Application of big data to behavioral and social sciences

Big data and biological scientists have worked collaboratively for decades, leading to important computational biology outcomes including the human genome project. Behavioral scientists are behind the curve.



Application of big data on behavioral & social sciences

Using big data can be an important alternative to advance social sciences in response to declines in survey participation and increased costs of data collection.

Application of big data on behavioral & social sciences

Training of public health and behavioral scientists in systems approaches and data science will be necessary to advance big data research in that domain.

CUNY SPH

Thank you!