

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

**Refining the Concept of Scientific Inference When Working With Big Data:
A Workshop**

*Sponsored by the National Institutes of Health Big Data to Knowledge Initiative
and the National Science Foundation*

June 8-9, 2016

*National Academy of Sciences Keck Center
500 Fifth St NW
Washington, DC
Room 100*

Workshop Objectives

Examine challenges in applying scientific inference to big data. To this end, the workshop will explore four key issues of scientific inference:

- Inference about causal discoveries driven by large observational data,
- Inference about discoveries from data on large networks,
- Inference about discoveries based on integration of diverse datasets, and
- Inference when regularization is used to simplify fitting of high-dimensional models.

In addressing these four issues, the workshop will:

- Bring together statisticians, data scientists, and domain researchers from different biomedical disciplines,
- Identify new methodologic developments that hold significant promise, and
- Highlight potential research program areas for the future.

Wednesday, June 8

8:30am Welcome and Overview

- Introductions from the Co-Chairs (*10 minutes*)
 - [Michael Daniels](#), University of Texas, Austin
 - [Alfred Hero](#), University of Michigan
- Perspectives from stakeholders (*30 minutes*)
 - [Phil Bourne](#), National Institutes of Health (*15 minutes*)
 - [Nandini Kannan](#), National Science Foundation, Division of Mathematical Sciences (*15 minutes*)
- Overview of the workshop (*30 minutes*)
 - [Michael Daniels](#), University of Texas at Austin

9:40am Break (*20 minutes*)

Committee on
Applied and
Theoretical
Statistics

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

10:00am Session I - Inference about discoveries based on integration of diverse datasets

Presenter (45 minutes)

- [Alfred Hero](#), University of Michigan, to speak about integrating and drawing inferences from multimodal data

Discussant (20 minutes)

- [Andrew Nobel](#), University of North Carolina, Chapel Hill

Q&A (20 minutes)

11:45am Lunch

12:45pm Session I, continued

Presenter (45 minutes)

- [Genevera Allen](#), Rice University, to speak about statistical methods using medical / health case studies

Discussant (20 minutes)

- [Jeffrey S. Morris](#), MD Anderson Cancer Center

Q&A (20 minutes)

2:10pm Break (20 minutes)

2:30pm Session II - Inference about causal discoveries driven by large observational data

Presenter (45 minutes)

- [Joe Hogan](#), Brown University, to speak about causal inference and decision making with health record data using a case study on HIV in Kenya

Discussant (20 minutes)

- [Elizabeth Stuart](#), Johns Hopkins University

Q&A (20 minutes)

3:55pm Break (20 minutes)

4:15pm Session II, continued

Presenter (45 minutes)

- [Sebastien Haneuse](#), Harvard University, to discuss comparative effectiveness research using electronic health records

Discussant (20 minutes)

- [Dylan Small](#), University of Pennsylvania

Q&A (20 minutes)

5:40pm Adjourn Day 1

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Thursday, June 9

8:30am Opening perspectives from stakeholders (10 minutes)

- [Chaitan Baru](#), National Science Foundation, Computer and Information Science and Engineering

8:40am Session III - Inference when regularization is used to simplify fitting of high-dimensional models

Presenter (40 minutes)

- [Daniela Witten](#), University of Washington, to discuss network reconstruction from high dimensional ordinary differential equations

Discussant (20 minutes)

- [Michael Kosorok](#), University of North Carolina at Chapel Hill

Q&A (20 minutes)

10:00am Break (20 minutes)

10:20am Session III, continued

Presenter (40 minutes)

- [Emery Brown](#), Massachusetts Institute of Technology, to speak about using different recording methods with high-dimensional time series

Discussant (20 minutes)

- [Xihong Lin](#), Harvard University

Q&A (20 minutes)

Technical/Methodological Presenter (40 minutes)

- [Jonathan Taylor](#), Stanford University

Q&A (10 minutes)

12:30pm Lunch

1:00pm Concluding Panel Discussion (2 hours)

Moderator: [Robert Kass](#), Carnegie Mellon University

Panelists:

- [Alfred Hero](#), University of Michigan
- [Bin Yu](#), University of California, Berkeley
- [Cosma Shalizi](#), Carnegie Mellon University
- [Andrew Nobel](#), University of North Carolina, Chapel Hill

3:00pm Adjourn

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Workshop Planning Committee

Co-Chairs

Michael J. Daniels, University of Texas at Austin

Alfred Hero, III, University of Michigan

Committee

Genevera Allen, Rice University and Baylor College of Medicine

Constantine Gatsonis, Brown University

Geoffrey Ginsburg, Duke University

Michael I. Jordan, University of California, Berkeley

Robert E. Kass, Carnegie Mellon University

Michael Kosorok, University of North Carolina at Chapel Hill

Roderick J.A. Little, University of Michigan

Jeffrey S. Morris, MD Anderson Cancer Center

Ronitt Rubinfeld, Massachusetts Institute of Technology

Lance Waller, Emory University

Staff

Michelle Schwalbe, mschwalbe@nas.edu

Rodney Howard, rhoward@nas.edu