Statistically Valid Inferences from Privacy Protected Data

Gary King

Institute for Quantitative Social Science
Harvard University

MIT Analytics Lab, 11/3/2022

GaryKing.org/privacy. Based on APSR/AJPS/PA articles with subsets of {Georgie Evans, Meg Schwenzfeier, Abhradeep Thakurta, Adam D. Smith}
VIEWPOINT: THE FUTURE

Through the Glass Lightly

A collection of scientists at the frontier were asked what they see in the future for science.*

Here are their views….

If you can look into the seeds of time,
And say which grain will grow and which will not,
Speak then to me, who neither beg nor fear
Your favors nor your hate.

Shakespeare, Macbeth, 1.3.58–61

There will be enormous inroads into human biology and human disease via genomics, gene therapy, and mouse knock-out models; a revolution in drug design by combinatorial chemistry; an understanding of the specificity of nerve connections and cognition; and the basic logic of development will be solved (if it is not solved already). New technologies will be developed for studying the structure, function, and dynamics of multiprotein ensembles—for example, the eukaryotic transcription complexes. New methodologies will be developed for studying the behavior of single, live cells in isolation or in the context of an embryo. This includes studying the activity of the cell itself as well as various subcellular structures.

Hal Weintraub
Fred Hutchinson Cancer Research Center
Seattle, Washington

Individuals at risk for diabetes, schizophrenia, obesity, and many other diseases. In many cases, disease will be either avoidable by modification of behavior or ameliorated by therapeutic intervention. For societies with socialized health care programs, the economic cost of screening will need to be balanced by the overall savings in disease reduction. If individuals refuse preventive treatment, screening is not cost-effective. For societies with private health care systems, the rich will become healthier and the poor sicker. In both systems, balancing the rights of individuals against the needs of society is going to be difficult.

Peter N. Goodfellow
Department of Genetics
University of Cambridge

Toxins, sunlight, and so forth. The output will be a color movie in which the embryo develops into a fetus, is born, and then grows into an adult, explicitly depicting body size and shape and hair, skin, and eye color. Eventually the DNA sequence base will be expanded to cover genes important for traits such as speech and musical ability; the mother will be able to hear the embryo—as an adult—speak or sing.

Harvey F. Lodish
Whitehead Institute for Biomedical Research
Cambridge, Massachusetts

The old phrase "you can't get blood from a turnip" may be proven incorrect, at least partially. Transgenic plants hold promise as biomanufacturing systems for a wide variety of human proteins, including those found in blood plasma. Serum albumin, for instance, has been shown to be expressed and processed correctly when the gene encoding it was introduced into plants. The missing element in this scenario is process technology, which will make it possible to do large-scale protein purification from plant tissues. Advances in high-level protein expression in specialized plant tissues (such as seeds, fruits, or tubers) coupled to engineering improvements in protein isolation may make this a reality.
Progress in Social Science

• What did 60 scientists forecast in 1995?

Fortunately, the social scientists in 1995 were wrong! We've seen spectacular progress, due to

• New data sources
  • surveys, end-of-period government stats, one-off studies of people, places, or events
  • text, images, video, social media, GIS, etc.

• New methods to analyze them

• Impact:
  • changed most Fortune 500 firms
  • established new industries
  • altered friendship networks, political campaigns, public health, legal analysis, policing, economics, sports, public policy, literature, etc., etc., etc.

Summary. Progress came from:

Novel data, novel methods
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists:

Fortunately, the social scientists in 1995 were wrong!

We've seen spectacular progress, due to

• New data sources
  • surveys, end-of-period government stats, one-off studies of people, places, or events
  • text, images, video, social media, GIS, etc.

• New methods to analyze them
• Impact:
  • changed most Fortune 500 firms
  • established new industries
  • altered friendship networks, political campaigns, public health, legal analysis, policing, economics, sports, public policy, literature, etc., etc., etc.

Summary. Progress came from:

Novel data, novel methods
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists:

• Social Scientists:

• New data sources
  • Surveys, end-of-period government stats, one-off studies of people, places, or events
• Now:
  • Text, images, video, social media, GIS, etc.
• New methods to analyze them
• Impact:
  • Changed most Fortune 500 firms
  • Established new industries
  • Altered friendship networks, political campaigns, public health, legal analysis, policing, economics, sports, public policy, literature, etc., etc., etc.
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists:
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this,
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

Fortunately, the social scientists in 1995 were wrong!

We've seen spectacular progress, due to

• New data sources
  • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
  • Now: text, images, video, social media, GIS, etc.

• New methods to analyze them
• Impact: changed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis, policing, economics, sports, public policy, literature, etc., etc., etc.

Summary. Progress came from: Novel data, novel methods
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!
• We’ve seen spectacular progress, due to

  • New data sources
  • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
  • Now: text, images, video, social media, GIS, etc.

  • New methods to analyze them

  • Impact: changed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis, policing, economics, sports, public policy, literature, etc., etc., etc.
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!
• We’ve seen spectacular progress, due to
  • New data sources
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!

• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!
• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    • Now: text, images, video, social media, GIS, etc.
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!
• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    • Now: text, images, video, social media, GIS, etc.
  • New methods to analyze them
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!

• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    • Now: text, images, video, social media, GIS, etc.
  • New methods to analyze them
  • Impact:
Progress in Social Science

- What did 60 scientists forecast in 1995?
  - Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  - Social Scientists: we study this, but soon will study that.

- Fortunately, the social scientists in 1995 were wrong!
- We’ve seen spectacular progress, due to
  - New data sources
    - Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    - Now: text, images, video, social media, GIS, etc.

- New methods to analyze them
- Impact: changed most Fortune 500 firms
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!
• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    • Now: text, images, video, social media, GIS, etc.
  • New methods to analyze them
  • Impact: changed most Fortune 500 firms; established new industries
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!
• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    • Now: text, images, video, social media, GIS, etc.
  • New methods to analyze them
  • Impact: changed most Fortune 500 firms; established new industries; altered friendship networks
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!
• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    • Now: text, images, video, social media, GIS, etc.
  • New methods to analyze them
  • Impact: changed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!

• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    • Now: text, images, video, social media, GIS, etc.
  • New methods to analyze them
  • Impact: changed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.
• Fortunately, the social scientists in 1995 were wrong!
• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    • Now: text, images, video, social media, GIS, etc.
  • New methods to analyze them
• Impact: changed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis
Progress in Social Science

- What did 60 scientists forecast in 1995?
  - Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  - Social Scientists: we study this, but soon will study that.
- Fortunately, the social scientists in 1995 were wrong!
- We’ve seen spectacular progress, due to
  - New data sources
    - Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    - Now: text, images, video, social media, GIS, etc.
  - New methods to analyze them
  - Impact: changed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis, policing
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!
• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    • Now: text, images, video, social media, GIS, etc.
  • New methods to analyze them
  • Impact: changed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis, policing, economics
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!

• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    • Now: text, images, video, social media, GIS, etc.
  • New methods to analyze them
  • Impact: changed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis, policing, economics, sports
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!
• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    • Now: text, images, video, social media, GIS, etc.

  • New methods to analyze them
  • Impact: changed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis, policing, economics, sports, public policy
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!
• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    • Now: text, images, video, social media, GIS, etc.
  • New methods to analyze them
  • Impact: changed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis, policing, economics, sports, public policy, literature,
Progress in Social Science

• What did 60 scientists forecast in 1995?
  • Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  • Social Scientists: we study this, but soon will study that.

• Fortunately, the social scientists in 1995 were wrong!

• We’ve seen spectacular progress, due to
  • New data sources
    • Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    • Now: text, images, video, social media, GIS, etc.
  • New methods to analyze them
  • Impact: changed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis, policing, economics, sports, public policy, literature, etc., etc., etc.
Progress in Social Science

- What did 60 scientists forecast in 1995?
  - Physical and natural scientists: breathtaking discoveries, inventions, engineering marvels, problems solved
  - Social Scientists: we study this, but soon will study that.

- Fortunately, the social scientists in 1995 were wrong!

- We’ve seen spectacular progress, due to
  - New data sources
    - Then: surveys, end-of-period government stats, one-off studies of people, places, or events
    - Now: text, images, video, social media, GIS, etc.
  - New methods to analyze them
  - Impact: changed most Fortune 500 firms; established new industries; altered friendship networks, political campaigns, public health, legal analysis, policing, economics, sports, public policy, literature, etc., etc., etc

- Summary. Progress came from: Novel data, novel methods
Present

Social scientists have more data than ever.

But a smaller % of data in the world than ever (about the people, groups, firms, countries we study).

Most is now locked up inside private companies and other orgs.

The central unresolved issue: Privacy (of customers, citizens, firms, etc.).

Future

We must liberate these datasets!

Academics, companies, governments, etc.: must get their privacy act together.

Goal today: data sharing without privacy violations.

How? Solving political problems technologically.
Progress in Social Science

• Present
Progress in Social Science

• Present

• Future
Progress in Social Science

• Present
  • Social scientists have more data than ever

• Future
Progress in Social Science

• Present
  • Social scientists have more data than ever
  • But a smaller % of data in the world than ever (about the people, groups, firms, countries we study)

• Future
Progress in Social Science

• Present
  • Social scientists have more data than ever
  • But a smaller % of data in the world than ever (about the people, groups, firms, countries we study)
  • Most is now locked up inside private companies and other orgs

• Future
Progress in Social Science

• Present
  • Social scientists have more data than ever
  • But a smaller % of data in the world than ever (about the people, groups, firms, countries we study)
  • Most is now locked up inside private companies and other orgs
  • The central unresolved issue: Privacy (of customers, citizens, firms, etc.)

• Future
Progress in Social Science

• Present
  • Social scientists have more data than ever
  • But a smaller % of data in the world than ever (about the people, groups, firms, countries we study)
  • Most is now locked up inside private companies and other orgs
  • The central unresolved issue: Privacy (of customers, citizens, firms, etc.)

• Future
  • We must liberate these datasets!
Progress in Social Science

• **Present**
  - Social scientists have more data than ever
  - But a smaller % of data in the world than ever (about the people, groups, firms, countries we study)
  - Most is now locked up inside private companies and other orgs
  - The central unresolved issue: Privacy (of customers, citizens, firms, etc.)

• **Future**
  - We must liberate these datasets!
  - Academics, companies, governments, etc.: must get their privacy act together
Progress in Social Science

• **Present**
  • Social scientists have more data than ever
  • But a smaller % of data in the world than ever (about the people, groups, firms, countries we study)
  • Most is now locked up inside private companies and other orgs
  • The central unresolved issue: Privacy (of customers, citizens, firms, etc.)

• **Future**
  • We must liberate these datasets!
  • Academics, companies, governments, etc.: must get their privacy act together
  • Goal today: data sharing without privacy violations
Progress in Social Science

- **Present**
  - Social scientists have more data than ever
  - But a smaller % of data in the world than ever (about the people, groups, firms, countries we study)
  - Most is now locked up inside private companies and other orgs
  - The central unresolved issue: Privacy (of customers, citizens, firms, etc.)

- **Future**
  - We must liberate these datasets!
  - Academics, companies, governments, etc.: must get their privacy act together
  - Goal today: data sharing without privacy violations
  - How? Solving political problems technologically
Solving Political Problems Technologically

Differential Privacy & Inferential Validity

A General Purpose, Statistically Valid DP Algorithm

The Algorithm in Practice
Convincing Facebook to Make Data Available

Gary visits Facebook to persuade them to make data available

In my hotel room packing, email arrives: “Hey what do we do about this?”

This was Cambridge Analytica. (The worst timed lobby effort in history!)

3 days later: “Could you do a study of the 2016 election?”

I’d love to, but I need 2 things & you’ll only give me 1:

- Complete access to data, people, etc. (like employees)
- No pre-publication approval (like NO employees ever)

We iterate, and I propose a 2-part solution

- Outside academics: send proposals, no company veto
- Trusted 3rd party: Commission at Social Science One signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges

Problem solved, without balancing agreements, announcements, funding, 30+ people assigned at Facebook

Just one issue: Facebook’s implementation plan was illegal!

New Problem: Sharing data without it leaving Facebook

Solving Political Problems Technologically
Convincing Facebook to Make Data Available
Solving a Political Problem Technologically (via “constitutional design”)

• Gary visits Facebook to persuade them to make data available
• In my hotel room packing, email arrives: “Hey what do we do about this?”
  This was Cambridge Analytica. (The worst timed lobby effort in history!)
• 3 days later: “Could you do a study of the 2016 election?”
• I’d love to, but I need 2 things & you’ll only give me 1:
  • Complete access to data, people, etc. (like employees)
  • No pre-publication approval (like NO employees ever)
• We iterate, and I propose a 2-part solution
  • Outside academics: send proposals, no company veto
  • Trusted 3rd party: Commission at Social Science One signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges
• Problem solved, without balancing agreements, announcements, funding, 30+ people assigned at Facebook
• Just one issue: Facebook’s implementation plan was illegal!
• New Problem: Sharing data without it leaving Facebook
Convincing Facebook to Make Data Available
Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to make data available

In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica. (The worst timed lobby effort in history!)

3 days later: “Could you do a study of the 2016 election?” I’d love to, but I need 2 things & you’ll only give me 1:

- Complete access to data, people, etc. (like employees)
- No pre-publication approval (like NO employees ever)

We iterate, and I propose a 2-part solution

- Outside academics: send proposals, no company veto
- Trusted 3rd party: Commission at Social Science One signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges

Problem solved, without balancing agreements, announcements, funding, 30+ people assigned at Facebook

Just one issue: Facebook’s implementation plan was illegal!

New Problem: Sharing data without it leaving Facebook
Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

• Gary visits Facebook to persuade them to make data available
• In my hotel room packing, email arrives: “Hey what do we do about this?”

...
Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

• Gary visits Facebook to persuade them to make data available
• In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica.

• Complete access to data, people, etc. (like employees)
• No pre-publication approval (like NO employees ever)

We iterate, and I propose a 2-part solution

• Outside academics: send proposals, no company veto
• Trusted 3rd party: Commission at Social Science One signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges

Problem solved, without balancing agreements, announcements, funding, 30+ people assigned at Facebook

• New Problem: Sharing data without it leaving Facebook
Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

• Gary visits Facebook to persuade them to make data available
• In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica. (The worst timed lobby effort in history!)
Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

• Gary visits Facebook to persuade them to make data available
• In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica. (The worst timed lobby effort in history!)
• 3 days later: “Could you do a study of the 2016 election?”
Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

• Gary visits Facebook to persuade them to make data available
• In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica. (The worst timed lobby effort in history!)
• 3 days later: “Could you do a study of the 2016 election?”
• I’d love to, but I need 2 things & you’ll only give me 1:
  • Complete access to data, people, etc. (like employees)
  • No pre-publication approval (like NO employees ever)

We iterate, and I propose a 2-part solution
  • Outside academics: send proposals, no company veto
  • Trusted 3rd party: Commission at Social Science One signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges

Problem solved, without balancing agreements, announcements, funding, 30+ people assigned at Facebook

Just one issue: Facebook’s implementation plan was illegal!

New Problem: Sharing data without it leaving Facebook
Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

• Gary visits Facebook to persuade them to make data available
• In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica. (The worst timed lobby effort in history!)
• 3 days later: “Could you do a study of the 2016 election?”
• I’d love to, but I need 2 things & you’ll only give me 1:
  • Complete access to data, people, etc. (like employees)
Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

• Gary visits Facebook to persuade them to make data available
• In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica. (The worst timed lobby effort in history!)
• 3 days later: “Could you do a study of the 2016 election?”
• I’d love to, but I need 2 things & you’ll only give me 1:
  • **Complete access** to data, people, etc. (like employees)
  • **No pre-publication approval** (like NO employees ever)

Problem solved, without balancing agreements, announcements, funding, 30+ people assigned at Facebook

New Problem: Sharing data without it leaving Facebook
Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

• Gary visits Facebook to persuade them to make data available
• In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica. (The worst timed lobby effort in history!)
• 3 days later: “Could you do a study of the 2016 election?”
• I’d love to, but I need 2 things & you’ll only give me 1:
  • Complete access to data, people, etc. (like employees)
  • No pre-publication approval (like NO employees ever)
• We iterate, and I propose a 2-part solution
Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

• Gary visits Facebook to persuade them to make data available
• In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica. (The worst timed lobby effort in history!)
• 3 days later: “Could you do a study of the 2016 election?”
• I’d love to, but I need 2 things & you’ll only give me 1:
  • Complete access to data, people, etc. (like employees)
  • No pre-publication approval (like NO employees ever)
• We iterate, and I propose a 2-part solution
  • Outside academics: send proposals, no company veto
Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to make data available
- In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica. (The worst timed lobby effort in history!)
- 3 days later: “Could you do a study of the 2016 election?”
- I’d love to, but I need 2 things & you’ll only give me 1:
  - Complete access to data, people, etc. (like employees)
  - No pre-publication approval (like NO employees ever)
- We iterate, and I propose a 2-part solution
  - Outside academics: send proposals, no company veto
  - Trusted 3rd party: Commission at Social Science One signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges
Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

- Gary visits Facebook to persuade them to make data available
- In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica. (The worst timed lobby effort in history!)
- 3 days later: “Could you do a study of the 2016 election?”
- I’d love to, but I need 2 things & you’ll only give me 1:
  - Complete access to data, people, etc. (like employees)
  - No pre-publication approval (like NO employees ever)
- We iterate, and I propose a 2-part solution
  - Outside academics: send proposals, no company veto
  - Trusted 3rd party: Commission at Social Science One signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges
- Problem solved, without balancing agreements, announcements, funding, 30+ people assigned at Facebook
Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

• Gary visits Facebook to persuade them to make data available
• In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica. (The worst timed lobby effort in history!)
• 3 days later: “Could you do a study of the 2016 election?”
• I’d love to, but I need 2 things & you’ll only give me 1:
  • Complete access to data, people, etc. (like employees)
  • No pre-publication approval (like NO employees ever)
• We iterate, and I propose a 2-part solution
  • Outside academics: send proposals, no company veto
  • Trusted 3rd party: Commission at Social Science One signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges
• Problem solved, without balancing ~ agreements, announcements, funding, 30+ people assigned at Facebook
• Just one issue:
Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via “constitutional design”)

• Gary visits Facebook to persuade them to make data available
• In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica. (The worst timed lobby effort in history!)
• 3 days later: “Could you do a study of the 2016 election?”
• I’d love to, but I need 2 things & you’ll only give me 1:
  • Complete access to data, people, etc. (like employees)
  • No pre-publication approval (like NO employees ever)
• We iterate, and I propose a 2-part solution
  • Outside academics: send proposals, no company veto
  • Trusted 3rd party: Commission at Social Science One signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges
• Problem solved, without balancing agreements, announcements, funding, 30+ people assigned at Facebook
• Just one issue: Facebook’s implementation plan was illegal!
Convincing Facebook to Make Data Available
Solving a Political Problem Technologically (via “constitutional design”)

• Gary visits Facebook to persuade them to make data available
• In my hotel room packing, email arrives: “Hey what do we do about this?” This was Cambridge Analytica. (The worst timed lobby effort in history!)
• 3 days later: “Could you do a study of the 2016 election?”
• I’d love to, but I need 2 things & you’ll only give me 1:
  • Complete access to data, people, etc. (like employees)
  • No pre-publication approval (like NO employees ever)
• We iterate, and I propose a 2-part solution
  • Outside academics: send proposals, no company veto
  • Trusted 3rd party: Commission at Social Science One signs NDAs, agree not to publish from the data, chooses datasets, makes final decisions; can report publicly if Facebook reneges
• Problem solved, without balancing agreements, announcements, funding, 30+ people assigned at Facebook
• Just one issue: Facebook’s implementation plan was illegal!
• New Problem: Sharing data without it leaving Facebook
Data Sharing Regime \(\sim\) Data Access Regime

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
- **Data Access Regime**: Trusted server holds data; researchers as adversaries, can run any method; noisy answer, a limited number of times

**Goal**: impossible to violate individual privacy; possible to discover population level patterns

\(\approx\) **differential privacy** (seems to satisfy regulators et al.)

**New Problem**: Most DP algorithms are statistically invalid!

unknown statistical properties (usually biased) - no uncertainty estimates
Data Sharing Regime $\leadsto$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

• Data Sharing Regime
  - I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!!)
  - Even trusting a researcher known to be trustworthy can fail

• Data Access Regime
  - Trusted server holds data; researchers as adversaries, can run any method; noisy answer, a limited number of times
  - Goal: impossible to violate individual privacy; & possible to discover population level patterns
  - $\approx$ differential privacy (seems to satisfy regulators et al.)

New Problem:
- Most DP algorithms are statistically invalid!
- Unknown statistical properties (usually biased)
- No uncertainty estimates
Data Sharing Regime $\leadsto$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)

- Venerable, but failing
- Increasing public concern with privacy
- Scholars discovered: de-identification doesn't work!
- Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
- Trusting researchers fails spectacularly at times (C.A.!)
  - Even trusting a researcher known to be trustworthy can fail
- Data Access Regime
  - Trusted server holds data; researchers as adversaries, can run any method; noisy answer, a limited number of times
- Goal: impossible to violate individual privacy; possible to discover population level patterns
  - $\approx$ differential privacy (seems to satisfy regulators et al.)
- New Problem: Most DP algorithms are statistically invalid!
  - unknown statistical properties (usually biased)
  - no uncertainty estimates
Data Sharing Regime $\sim$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)

- **Data Access Regime**
Data Sharing Regime \xrightarrow{\sim} Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but **failing**

- **Data Access Regime**
Data Sharing Regime $\Rightarrow$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy

- **Data Access Regime**
Data Sharing Regime $\leadsto$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but **failing**
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!

- **Data Access Regime**
Data Sharing Regime $\sim$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does

- **Data Access Regime**
Data Sharing Regime $\sim$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation,

- **Data Access Regime**

≈ differential privacy (seems to satisfy regulators et al.)

New Problem: Most DP algorithms are statistically invalid!

unknown statistical properties (usually biased)

no uncertainty estimates
Data Sharing Regime $\sim$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing,

- **Data Access Regime**
Data Sharing Regime $\Rightarrow$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but **failing**
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms,

- **Data Access Regime**
Data Sharing Regime $\rightsquigarrow$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but *failing*
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements,

- **Data Access Regime**

Solving Political Problems Technologically
Data Sharing Regime $\rightsquigarrow$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime:** I give you data (maybe you sign DUA)
  - Venerable, but **failing**
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing,

- **Data Access Regime**
Data Sharing Regime $\sim$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models,

- **Data Access Regime**
Data Sharing Regime $\leadsto$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.

- **Data Access Regime**
Data Sharing Regime $\leadsto$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime:** I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!!)

- **Data Access Regime**
Data Sharing Regime $\sim$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!!)
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
Data Sharing Regime $\sim$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but **failing**
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!)  
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
  - Trusted server holds data;
Data Sharing Regime $\rightsquigarrow$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but **failing**
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!)  
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
  - Trusted server holds data; researchers as adversaries
Data Sharing Regime $\leadsto$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!)  
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
  - Trusted server holds data; researchers as adversaries, can run any method $\leadsto$ noisy answer,
Data Sharing Regime $\rightarrow$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A. !)
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
  - Trusted server holds data; researchers as adversaries, can run any method $\rightarrow$ noisy answer, a limited number of times

≈ differential privacy (seems to satisfy regulators et al.)

New Problem: Most DP algorithms are statistically invalid!

unknown statistical properties (usually biased)

no uncertainty estimates
Data Sharing Regime $\leadsto$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but **failing**
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!)  
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
  - Trusted server holds data; researchers as adversaries, can run any method $\leadsto$ noisy answer, a limited number of times
  - **Goal:**
Data Sharing Regime $\leadsto$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but **failing**
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!)  
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
  - Trusted server holds data; researchers as adversaries, can run any method $\leadsto$ noisy answer, a limited number of times
  - **Goal**: **impossible** to violate individual privacy
Data Sharing Regime $\rightsquigarrow$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but **failing**
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!)  
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
  - Trusted server holds data; researchers as adversaries, can run any method $\rightsquigarrow$ noisy answer, a limited number of times
  - **Goal**: impossible to violate individual privacy; & possible to discover population level patterns
Data Sharing Regime $\leadsto$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!)  
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
  - Trusted server holds data; researchers as adversaries, can run any method $\leadsto$ noisy answer, a limited number of times
  - **Goal**: impossible to violate individual privacy; & possible to discover population level patterns
  - $\approx$ differential privacy
Data Sharing Regime $\leadsto$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!) 
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
  - Trusted server holds data; researchers as adversaries, can run any method $\leadsto$ noisy answer, a limited number of times
  - **Goal**: impossible to violate individual privacy; & possible to discover population level patterns
  - $\approx$ **differential privacy** (seems to satisfy regulators et al.)
Data Sharing Regime $\sim$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!)  
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
  - Trusted server holds data; researchers as adversaries, can run any method $\sim$ noisy answer, a limited number of times
  - **Goal**: impossible to violate individual privacy; & possible to discover population level patterns
  - $\approx$ differential privacy (seems to satisfy regulators et al.)
  - **New Problem:**
Data Sharing Regime $\leadsto$ Data Access Regime
Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A. !)
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
  - Trusted server holds data; researchers as adversaries, can run any method $\leadsto$ noisy answer, a limited number of times
  - **Goal**: impossible to violate individual privacy; & possible to discover population level patterns
  - $\approx$ differential privacy (seems to satisfy regulators et al.)
  - **New Problem**: Most DP algorithms are statistically invalid!
Data Sharing Regime $\sim$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but *failing*
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!)  
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
  - Trusted server holds data; researchers as adversaries, can run any method $\sim$ noisy answer, a limited number of times
  - **Goal**: impossible to violate individual privacy; & possible to discover population level patterns
  - $\approx$ differential privacy (seems to satisfy regulators et al.)
  - **New Problem**: Most DP algorithms are statistically invalid!
    - *unknown* statistical properties (usually *biased*)
Data Sharing Regime $\leadsto$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

- **Data Sharing Regime**: I give you data (maybe you sign DUA)
  - Venerable, but failing
  - Increasing public concern with privacy
  - Scholars discovered: de-identification doesn’t work!
  - Nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!) 
  - Even trusting a researcher known to be trustworthy can fail

- **Data Access Regime**
  - Trusted server holds data; researchers as adversaries, can run any method $\leadsto$ noisy answer, a limited number of times
  - **Goal**: impossible to violate individual privacy; & possible to discover population level patterns
  - $\approx$ differential privacy (seems to satisfy regulators et al.)
  - **New Problem**: Most DP algorithms are statistically invalid!
    - unknown statistical properties (usually biased)
    - no uncertainty estimates
Solving Political Problems Technologically

Differential Privacy & Inferential Validity

A General Purpose, Statistically Valid DP Algorithm

The Algorithm in Practice
Theories of Inference: Statistics vs. CS

<table>
<thead>
<tr>
<th>Name</th>
<th>Quantity of Interest</th>
<th>Mean Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rocio</td>
<td>3</td>
<td>122</td>
</tr>
<tr>
<td>John</td>
<td>3</td>
<td>76</td>
</tr>
<tr>
<td>Marc</td>
<td>3</td>
<td>145</td>
</tr>
<tr>
<td>Brandon</td>
<td>3</td>
<td>96</td>
</tr>
<tr>
<td>Yu Xie</td>
<td>3</td>
<td>86</td>
</tr>
<tr>
<td>Gleason</td>
<td>3</td>
<td>127</td>
</tr>
<tr>
<td>Saad</td>
<td>3</td>
<td>72</td>
</tr>
<tr>
<td>Leonard</td>
<td>3</td>
<td>132</td>
</tr>
<tr>
<td>Zhou</td>
<td>3</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>134</td>
</tr>
</tbody>
</table>
Theories of Inference: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Quantity of Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean income: $48</td>
</tr>
</tbody>
</table>

| Rocío       | 3                     |
| John        | 3                     |
| Marc        | 3                     |
| Brandon     | 3                     |
| Yu Xie      | 3                     |
| Gleason     | 3                     |
| Saad        | 3                     |
| Leonard     | 3                     |
| Kristopher  | 3                     |
| Zhou        | 3                     |

$\vdash$ Privacy $= \text{dp}$
## Theories of Inference: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>rocío</td>
<td>✓</td>
</tr>
<tr>
<td>John</td>
<td>✓</td>
</tr>
<tr>
<td>marc</td>
<td>✓</td>
</tr>
<tr>
<td>Brandon</td>
<td>✓</td>
</tr>
<tr>
<td>Yu Xie</td>
<td>✓</td>
</tr>
<tr>
<td>Gleason</td>
<td>✓</td>
</tr>
<tr>
<td>Saad</td>
<td>✓</td>
</tr>
<tr>
<td>Leonard</td>
<td>✓</td>
</tr>
<tr>
<td>Kristopher</td>
<td>✓</td>
</tr>
<tr>
<td>Zhou</td>
<td>✓</td>
</tr>
</tbody>
</table>

Mean income: $48
### Theories of Inference: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>⋮</td>
<td>⋱</td>
<td>?</td>
</tr>
<tr>
<td>Rocío</td>
<td>✓</td>
<td>122</td>
</tr>
<tr>
<td>John</td>
<td>✓</td>
<td>76</td>
</tr>
<tr>
<td>Marc</td>
<td>✓</td>
<td>145</td>
</tr>
<tr>
<td>Brandon</td>
<td>✓</td>
<td>96</td>
</tr>
<tr>
<td>Yu Xie</td>
<td>✓</td>
<td>86</td>
</tr>
<tr>
<td>Gleason</td>
<td>✓</td>
<td>127</td>
</tr>
<tr>
<td>Saad</td>
<td>✓</td>
<td>72</td>
</tr>
<tr>
<td>Leonard</td>
<td>✓</td>
<td>132</td>
</tr>
<tr>
<td>Kristopher</td>
<td>✓</td>
<td>95</td>
</tr>
<tr>
<td>Zhou</td>
<td>✓</td>
<td>134</td>
</tr>
</tbody>
</table>

**Mean income:**

- Classical Inference: $48
- $108

---

**Quantity of Interest**

- Usually no direct relevance
# Theories of Inference: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>✗</td>
<td>?</td>
</tr>
<tr>
<td>Rocío</td>
<td>✓</td>
<td>122</td>
</tr>
<tr>
<td>John</td>
<td>✓</td>
<td>76</td>
</tr>
<tr>
<td>Marc</td>
<td>✓</td>
<td>145</td>
</tr>
<tr>
<td>Brandon</td>
<td>✓</td>
<td>96</td>
</tr>
<tr>
<td>Yu Xie</td>
<td>✓</td>
<td>86</td>
</tr>
<tr>
<td>Gleason</td>
<td>✓</td>
<td>127</td>
</tr>
<tr>
<td>Saad</td>
<td>✓</td>
<td>72</td>
</tr>
<tr>
<td>Leonard</td>
<td>✓</td>
<td>132</td>
</tr>
<tr>
<td>Kristopher</td>
<td>✓</td>
<td>95</td>
</tr>
<tr>
<td>Zhou</td>
<td>✓</td>
<td>134</td>
</tr>
</tbody>
</table>

Mean income: $48, Classical Inference: $108

**Quantity of Interest**

**Usually no direct relevance**
Theories of Inference: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
<th>$</th>
<th>+Privacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>:</td>
<td>X</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Rocío</td>
<td>✓</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>✓</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>Marc</td>
<td>✓</td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>Brandon</td>
<td>✓</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>Yu Xie</td>
<td>✓</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>Gleason</td>
<td>✓</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>Saad</td>
<td>✓</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Leonard</td>
<td>✓</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>Kristopher</td>
<td>✓</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>Zhou</td>
<td>✓</td>
<td>134</td>
<td></td>
</tr>
</tbody>
</table>

Mean income:

<table>
<thead>
<tr>
<th>Quantity of Interest</th>
<th>Classical Inference</th>
<th>Usually no direct relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$48</td>
<td>$108</td>
<td></td>
</tr>
</tbody>
</table>
## Theories of Inference: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
<th>$</th>
<th>+Privacy</th>
<th>=dp$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>✗</td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rocío</td>
<td>✓</td>
<td>122</td>
<td></td>
<td>85</td>
</tr>
<tr>
<td>John</td>
<td>✓</td>
<td>76</td>
<td></td>
<td>103</td>
</tr>
<tr>
<td>Marc</td>
<td>✓</td>
<td>145</td>
<td></td>
<td>75</td>
</tr>
<tr>
<td>Brandon</td>
<td>✓</td>
<td>96</td>
<td></td>
<td>113</td>
</tr>
<tr>
<td>Yu Xie</td>
<td>✓</td>
<td>86</td>
<td></td>
<td>125</td>
</tr>
<tr>
<td>Gleason</td>
<td>✓</td>
<td>127</td>
<td></td>
<td>97</td>
</tr>
<tr>
<td>Saad</td>
<td>✓</td>
<td>72</td>
<td></td>
<td>101</td>
</tr>
<tr>
<td>Leonard</td>
<td>✓</td>
<td>132</td>
<td></td>
<td>128</td>
</tr>
<tr>
<td>Kristopher</td>
<td>✓</td>
<td>95</td>
<td></td>
<td>83</td>
</tr>
<tr>
<td>Zhou</td>
<td>✓</td>
<td>134</td>
<td></td>
<td>201</td>
</tr>
</tbody>
</table>

**Mean income:**

- Classical Inference: $48
- Query-Response: $108
- Privacy Protected: $111

**Quantity of Interest:** Usually no direct relevance

**Noise & Censoring:**

- No direct relevance
### Theories of Inference: Statistics vs. CS

<table>
<thead>
<tr>
<th>Population</th>
<th>Sample</th>
<th>$</th>
<th>+Privacy</th>
<th>=dp$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*</td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rocío</td>
<td>✔</td>
<td>122</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>✔</td>
<td>76</td>
<td>103</td>
<td></td>
</tr>
<tr>
<td>Marc</td>
<td>✔</td>
<td>145</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>Brandon</td>
<td>✔</td>
<td>96</td>
<td>113</td>
<td></td>
</tr>
<tr>
<td>Yu Xie</td>
<td>✔</td>
<td>86</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>Gleason</td>
<td>✔</td>
<td>127</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>Saad</td>
<td>✔</td>
<td>72</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>Leonard</td>
<td>✔</td>
<td>132</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>Kristopher</td>
<td>✔</td>
<td>95</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>Zhou</td>
<td>✔</td>
<td>134</td>
<td>201</td>
<td></td>
</tr>
</tbody>
</table>

Mean income: $48 → $108 → $111

Statistically Valid Inferences from Privacy Protected Data

---

Differential Privacy & Inferential Validity
Protecting Survey Data

Differential Privacy & Inferential Validity
Differential Privacy and its Inferential Challenges

Classical Statistics: Apply statistic $s(D)$

DP Mechanism: $M(s, D)$, with noise & censoring

Essential components of ensuring privacy

Fundamental problems for statistical inference

The DP Standard (simplifying)

Including ($D$) or excluding ($D'$) you doesn't change conclusions

$$\Pr[M(s, D) = m] \in 1 \pm \epsilon$$

for all $D, D', m$

Examples all proven to protect the biggest possible outlier

$$M(\text{mean}, D) = \frac{1}{n} \sum_{i=1}^{n} c(y_i, \Lambda) + N(0, \Lambda \frac{8}{n} \epsilon)$$

($\Lambda, n, \epsilon$ known)

Or: mess with gradients, $X'_i, X_i$, data, QOIs, etc.

Statistical properties: usually biased, no uncertainty estimates
Differential Privacy and its Inferential Challenges

• Estimators
Differential Privacy and its Inferential Challenges

• **Estimators**
  
  • **Classical Statistics:** Apply statistic $s$ to dataset $D$, $s(D)$
Differential Privacy and its Inferential Challenges

- **Estimators**
  - **Classical Statistics:** Apply statistic $s$ to dataset $D$, $s(D)$
  - **DP Mechanism:** $M(s, D)$, with noise & censoring

Examples all proven to protect the biggest possible outlier

$\Pr[M(s, D) = m] \approx 1 \pm \epsilon$

Including ($D$) or excluding ($D'$) you doesn't change conclusions

Statistical properties: usually biased, no uncertainty estimates
Differential Privacy and its Inferential Challenges

• Estimators
  • Classical Statistics: Apply statistic $s$ to dataset $D$, $s(D)$
  • DP Mechanism: $M(s, D)$, with noise & censoring
    • Essential components of ensuring privacy
Differential Privacy and its Inferential Challenges

- **Estimators**
  - **Classical Statistics:** Apply statistic $s$ to dataset $D$, $s(D)$
  - **DP Mechanism:** $M(s, D)$, with noise & censoring
    - Essential components of ensuring privacy
    - Fundamental problems for statistical inference

Differential Privacy & Inferential Validity
Differential Privacy and its Inferential Challenges

• Estimators
  • Classical Statistics: Apply statistic $s$ to dataset $D$, $s(D)$
  • DP Mechanism: $M(s, D)$, with noise & censoring
    • Essential components of ensuring privacy
    • Fundamental problems for statistical inference

• The DP Standard (simplifying)
Differential Privacy and its Inferential Challenges

• Estimators
  • Classical Statistics: Apply statistic $s$ to dataset $D$, $s(D)$
  • DP Mechanism: $M(s, D)$, with noise & censoring
    • Essential components of ensuring privacy
    • Fundamental problems for statistical inference

• The DP Standard (simplifying)
  • Including ($D$) or excluding ($D'$) you doesn’t change conclusions

\[
\frac{\Pr[M(s, D) = m]}{\Pr[M(s, D') = m]} \in 1 \pm \epsilon
\]

for all $D, D', m$
Differential Privacy and its Inferential Challenges

• **Estimators**
  - **Classical Statistics:** Apply statistic $s$ to dataset $D$, $s(D)$
  - **DP Mechanism:** $M(s, D)$, with noise & censoring
    - Essential components of ensuring privacy
    - Fundamental problems for statistical inference

• **The DP Standard (simplifying)**
  - Including $(D)$ or excluding $(D')$ you doesn’t change conclusions
    \[
    \frac{\Pr[M(s, D) = m]}{\Pr[M(s, D') = m]} \in 1 \pm \epsilon
    \]
    for all $D, D', m$
  - **Examples** all proven to protect the biggest possible outlier
Differential Privacy and its Inferential Challenges

- **Estimators**
  - **Classical Statistics:** Apply statistic $s$ to dataset $D$, $s(D)$
  - **DP Mechanism:** $M(s, D)$, with noise & censoring
    - Essential components of ensuring privacy
    - Fundamental problems for statistical inference

- **The DP Standard (simplifying)**
  - Including $(D)$ or excluding $(D')$ you doesn’t change conclusions
    \[
    \frac{\Pr[M(s, D) = m]}{\Pr[M(s, D') = m]} \in 1 \pm \epsilon
    \]
    for all $D, D', m$

- **Examples** all proven to protect the biggest possible outlier
  - $M(\text{mean}, D) = \frac{1}{n} \sum_{i=1}^{n} c(y_i, \Lambda) + N \left(0, \frac{8\Lambda}{n\epsilon}\right)$  
    ($\Lambda, n, \epsilon$ known)
Differential Privacy and its Inferential Challenges

• Estimators
  • Classical Statistics: Apply statistic $s$ to dataset $D$, $s(D)$
  • DP Mechanism: $M(s, D)$, with noise & censoring
    • Essential components of ensuring privacy
    • Fundamental problems for statistical inference

• The DP Standard (simplifying)
  • Including $(D)$ or excluding $(D')$ you doesn’t change conclusions

$$\frac{Pr[M(s, D) = m]}{Pr[M(s, D') = m]} \in 1 \pm \epsilon$$

for all $D, D', m$

• Examples all proven to protect the biggest possible outlier
  • $M(\text{mean}, D) = \frac{1}{n} \sum_{i=1}^{n} c(y_i, \Lambda) + N(0, \frac{8\Lambda}{n\epsilon})$ (\(\Lambda, n, \epsilon \) known)
  • Or: mess with gradients, $X'_iX_i$, data, QOIs, etc.
Differential Privacy and its Inferential Challenges

- **Estimators**
  - **Classical Statistics:** Apply statistic \( s \) to dataset \( D \), \( s(D) \)
  - **DP Mechanism:** \( M(s, D) \), with noise & censoring
    - Essential components of ensuring privacy
    - Fundamental problems for statistical inference

- **The DP Standard (simplifying)**
  - Including \( (D) \) or excluding \( (D') \) you doesn’t change conclusions
    \[
    \frac{\Pr[M(s, D) = m]}{\Pr[M(s, D') = m]} \in 1 \pm \epsilon
    \]
    for all \( D, D', m \)
  - **Examples** all proven to protect the biggest possible outlier
    - \( M(\text{mean}, D) = \frac{1}{n} \sum_{i=1}^{n} c(y_i, \Lambda) + N \left( 0, \frac{8\Lambda}{n\epsilon} \right) \) \( (\Lambda, n, \epsilon \text{ known}) \)
    - Or: mess with gradients, \( X'_i X_i \), data, QOIs, etc.
  - **Statistical properties:** usually biased, no uncertainty estimates
Solving Political Problems Technologically

Differential Privacy & Inferential Validity

A General Purpose, Statistically Valid DP Algorithm

The Algorithm in Practice
A Differentially Private Estimator

Private data \( \mathcal{D} \)

Partition \( \mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{D}_5 \)

Bag of little bootstraps

Estimator \( \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4, \hat{\theta}_5 \)

Censor

Average Noise \( \hat{\theta}_{dp} = \frac{1}{P} \sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N(0, \Lambda^2 \epsilon) \)

Bias Correction (and variance estimation)
A Differentially Private Estimator

Private data

A General Purpose, Statistically Valid DP Algorithm
A Differentially Private Estimator

Private data

Partition

A General Purpose, Statistically Valid DP Algorithm
A Differentially Private Estimator

Private data
Partition
Bag of little bootstraps
A Differentially Private Estimator

Private data
Partition
Bag of little bootstraps
Estimator
A Differentially Private Estimator

Private data

Partition

Bag of little bootstraps

Estimator

Censor

A General Purpose, Statistically Valid DP Algorithm
A Differentially Private Estimator

Private data
Partition
Bag of little bootstraps
Estimator
Censor
Average

A General Purpose, Statistically Valid DP Algorithm
A Differentially Private Estimator

Private data
Partition
Bag of little bootstraps
Estimator
Censor
Average
Noise

A General Purpose, Statistically Valid DP Algorithm
A Differentially Private Estimator

Private data

Partition

Bag of little bootstraps

Estimator

Censor

Average

Noise

Bias Correction

A General Purpose, Statistically Valid DP Algorithm
A Differentially Private Estimator

Private data
Partition
Bag of little bootstraps
Estimator
Censor
Average
Noise
Bias Correction (& variance estimation)

A General Purpose, Statistically Valid DP Algorithm
A Differentially Private Estimator

Private data
Partition
Bag of little bootstraps
Estimator
Censor
Average
Noise
Bias Correction (& variance estimation)

A General Purpose, Statistically Valid DP Algorithm
A Differentially Private Estimator

Private data
Partition
Bag of little bootstraps
Estimator
Censor
Average
Noise
Bias Correction
(& variance estimation)

A General Purpose, Statistically Valid DP Algorithm
A Differentially Private Estimator

Private data

Partition

Bag of little bootstraps

Estimator

Censor

Average

Noise

Bias Correction

A General Purpose, Statistically Valid DP Algorithm
A Differentially Private Estimator

Private data

Partition

Bag of little bootstraps

Estimator

Censor

Average

Noise

Bias Correction
(& variance estimation)
A Differentially Private Estimator

Private data
Partition
Bag of little bootstraps
Estimator
Censor
Average
Noise
Bias Correction (& variance estimation)

\[ \hat{\theta}^{dp} = \frac{1}{P} \sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N\left(0, \frac{8\Lambda}{P\epsilon}\right) \]
Bias Correction of: \[
\hat{\theta}^{dp} = \frac{1}{P} \sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N \left( 0, \frac{8\Lambda}{P\epsilon} \right) \quad (\Lambda, P, \epsilon \text{ known})
\]
Bias Correction of:
\[ \hat{\theta}^{dp} = \frac{1}{P} \sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N \left( 0, \frac{8\Lambda}{P\epsilon} \right) \]

\( \Lambda, P, \epsilon \) known

\[ \hat{\theta}_p \sim N(\theta, \sigma^2) \]

Uncensored

\( \theta \)

Goal
Bias Correction of: 

\[ \hat{\theta}^{dp} = \frac{1}{P} \sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N \left( 0, \frac{8\Lambda}{P\epsilon} \right) \]  

(\Lambda, P, \epsilon \text{ known})

Uncensored

\[ \hat{\theta}_p \sim N(\theta, \sigma^2) \]

Censored distribution

Goal
Bias Correction of: 
\[
\hat{\theta}^{dp} = \frac{1}{P} \sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N\left(0, \frac{8\Lambda}{P\epsilon}\right) 
\]
(\Lambda, P, \epsilon \text{ known})

\(\hat{\theta}_p \sim N(\theta, \sigma^2)\)

Uncensored

Censored distribution

\(\alpha = \int_{\Lambda}^{\infty} N(t \mid \theta, \sigma^2) dt\)

Goal

\(\theta_c \quad \theta \quad \Lambda\)
Bias Correction of:  

\[ \hat{\theta}^{dp} = \frac{1}{P} \sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N \left( 0, \frac{8\Lambda}{P\epsilon} \right) \]  

(\Lambda, P, \epsilon \text{ known})

\[ \hat{\theta}_p \sim N(\theta, \sigma^2) \]  

Uncensored

\[ \alpha = \int_{\Lambda}^{\infty} N(t \mid \theta, \sigma^2) dt \]

\[ \theta_c = (1 - \alpha)\theta_T + \alpha\Lambda \]  

Goal
Bias Correction of: 

\[ \hat{\theta}^{dp} = \frac{1}{P} \sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N \left( 0, \frac{8\Lambda}{P \epsilon} \right) \]  

(\Lambda, P, \epsilon \text{ known})

Uncensored

\[ \hat{\theta}_p \sim N(\theta, \sigma^2) \]

Censored distribution

\[ \alpha = \int_{\Lambda}^{\infty} N(t \mid \theta, \sigma^2) dt \]

\[ \theta_c = (1 - \alpha)\theta_T + \alpha\Lambda \]

Goal

Equations: 2
Bias Correction of: \[ \hat{\theta}^{dp} = \frac{1}{P} \sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N\left(0, \frac{8\Lambda}{P\epsilon}\right) \] (\(\Lambda, P, \epsilon\) known)

\[ \hat{\theta}_p \sim N(\theta, \sigma^2) \]

Uncensored

\[ \theta_c = (1 - \alpha)\theta_T + \alpha\Lambda \]

Goal

Equations: 2

Unknowns: \(\theta, \sigma^2, \alpha, \theta_c\)
Bias Correction of:  
\[ \hat{\theta}^{dp} = \frac{1}{P} \sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N\left(0, \frac{8\Lambda}{P\epsilon}\right) \]  
\((\Lambda, P, \epsilon \text{ known})\)

Uncensored
\[ \hat{\theta}_p \sim N(\theta, \sigma^2) \]

Censored distribution
\[ \alpha = \int_{\Lambda}^{\infty} N(t \mid \theta, \sigma^2) dt \]

\[ \theta_c = (1 - \alpha)\theta_T + \alpha\Lambda \]

Disclose: \[ \hat{\theta}^{dp} \]

Equations: 2
Unknowns: \( \theta, \sigma^2, \alpha, \chi \)
Bias Correction of: 

$$\hat{\theta}_{dp} = \frac{1}{P} \sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N\left(0, \frac{8\Lambda}{P\epsilon}\right)$$  

($\Lambda, P, \epsilon$ known)

Uncensored 

$$\hat{\theta}_p \sim N(\theta, \sigma^2)$$

Censored distribution

$$\theta_c = (1 - \alpha)\theta_T + \alpha\Lambda$$

Goal

Disclose: $\hat{\theta}_{dp}, \hat{\alpha}_{dp}$

$$\alpha = \int_{\Lambda}^{\infty} N(t \mid \theta, \sigma^2)dt$$

Equations: 2

Unknowns: $\theta, \sigma^2, x, x_c$
Variance Estimation

Simulate estimates via standard (Clarify) procedures:

\[ \hat{\theta}_{dp}, \hat{\alpha}_{dp} \sim N(\begin{bmatrix} \hat{\theta}_{dp} \\ \hat{\alpha}_{dp} \end{bmatrix}, \begin{bmatrix} \hat{V}(\hat{\theta}_{dp}) & \hat{\text{Cov}}(\hat{\alpha}_{dp}, \hat{\theta}_{dp}) \\ \hat{\text{Cov}}(\hat{\alpha}_{dp}, \hat{\theta}_{dp}) & \hat{V}(\hat{\alpha}_{dp}) \end{bmatrix}) \)

Functions of disclosed params

Bias correct simulated params:

\{ \tilde{\theta}_{dp}, \hat{\sigma}^2_{dp} \} = \text{BiasCorrect}[\hat{\theta}_{dp}, \hat{\alpha}_{dp}]

Standard error:

Standard deviation of \( \tilde{\theta}_{dp} \) over simulations

Bias correction:

reduces bias and variance
Variance Estimation

- **Simulate estimates** via standard (Clarify) procedures:

\[
\hat{\theta}^{dp}, \hat{\alpha}^{dp} \sim N\left(\begin{bmatrix} \hat{\theta}^{dp} \\ \hat{\alpha}^{dp} \end{bmatrix}, \begin{bmatrix} \hat{V}(\hat{\theta}^{dp}) & \hat{\text{Cov}}(\hat{\alpha}^{dp}, \hat{\theta}^{dp}) \\ \hat{\text{Cov}}(\hat{\alpha}^{dp}, \hat{\theta}^{dp}) & \hat{V}(\hat{\alpha}^{dp}) \end{bmatrix}\right)
\]
Variance Estimation

- **Simulate estimates** via standard (Clarify) procedures:

\[
\hat{\theta}^{dp}, \hat{\alpha}^{dp} \sim N\left(\begin{bmatrix} \hat{\theta}^{dp} \\ \hat{\alpha}^{dp} \end{bmatrix}, \begin{bmatrix} \hat{V}(\hat{\theta}^{dp}) & \hat{\text{Cov}}(\hat{\alpha}^{dp}, \hat{\theta}^{dp}) \\ \hat{\text{Cov}}(\hat{\alpha}^{dp}, \hat{\theta}^{dp}) & \hat{V}(\hat{\alpha}^{dp}) \end{bmatrix}\right)
\]

Functions of disclosed params
Variance Estimation

- **Simulate estimates** via standard (Clarify) procedures:

\[
\hat{\theta}^{dp}, \hat{\alpha}^{dp} \sim N\left(\begin{bmatrix} \hat{\theta}^{dp} \\ \hat{\alpha}^{dp} \end{bmatrix}, \begin{bmatrix} \hat{V}(\hat{\theta}^{dp}) & \hat{\text{Cov}}(\hat{\alpha}^{dp}, \hat{\theta}^{dp}) \\ \hat{\text{Cov}}(\hat{\alpha}^{dp}, \hat{\theta}^{dp}) & \hat{V}(\hat{\alpha}^{dp}) \end{bmatrix}\right)
\]

- **Bias correct simulated params:**

\[
\{\tilde{\theta}^{dp}, \tilde{\sigma}^2_{dp}\} = \text{BiasCorrect}\left[\hat{\theta}^{dp}, \hat{\alpha}^{dp}\right]
\]
Variance Estimation

- **Simulate estimates** via standard (Clarify) procedures:

\[
\hat{\theta}^{dp}, \hat{\alpha}^{dp} \sim N\left(\begin{bmatrix} \hat{\theta}^{dp} \\ \hat{\alpha}^{dp} \end{bmatrix}, \begin{bmatrix} \hat{V}(\hat{\theta}^{dp}) & \hat{\text{Cov}}(\hat{\alpha}^{dp}, \hat{\theta}^{dp}) \\ \hat{\text{Cov}}(\hat{\alpha}^{dp}, \hat{\theta}^{dp}) & \hat{V}(\hat{\alpha}^{dp}) \end{bmatrix}\right)
\]

- **Bias correct simulated params:**

\[\{\tilde{\theta}^{dp}, \tilde{\sigma}^{2}_{dp}\} = \text{BiasCorrect} \left[ \hat{\theta}^{dp}, \hat{\alpha}^{dp} \right]\]

- **Standard error:** Standard deviation of \(\tilde{\theta}^{dp}\) over simulations
Variance Estimation

- **Simulate estimates** via standard (Clarify) procedures:

\[
\hat{\theta}_{dp}, \hat{\alpha}_{dp} \sim N\left(\begin{bmatrix} \hat{\theta}_{dp} \\ \hat{\alpha}_{dp} \end{bmatrix}, \begin{bmatrix} \hat{V}(\hat{\theta}_{dp}) & \hat{\text{Cov}}(\hat{\alpha}_{dp}, \hat{\theta}_{dp}) \\ \hat{\text{Cov}}(\hat{\alpha}_{dp}, \hat{\theta}_{dp}) & \hat{V}(\hat{\alpha}_{dp}) \end{bmatrix}\right)
\]

- **Bias correct simulated params:**

\[\{\tilde{\theta}_{dp}, \tilde{\sigma}^2_{dp}\} = \text{BiasCorrect}\left[\hat{\theta}_{dp}, \hat{\alpha}_{dp}\right]\]

- **Standard error:** Standard deviation of \(\tilde{\theta}_{dp}\) over simulations

- **Bias correction:** reduces bias *and* variance
Solving Political Problems Technologically

Differential Privacy & Inferential Validity

A General Purpose, Statistically Valid DP Algorithm

The Algorithm in Practice
Simulations: Finite Sample Evaluation
Simulations: Finite Sample Evaluation

The Algorithm in Practice
Simulations: Finite Sample Evaluation

The Algorithm in Practice
Simulations: Finite Sample Evaluation

The Algorithm in Practice
Simulations: Finite Sample Evaluation

\[ S E^d_p \theta \sim d_p \]

\[ S E^\wedge_d_p \theta \]

\[ S E^\sim_d_p \theta \]

The Algorithm in Practice
Simulations: Finite Sample Evaluation

The Algorithm in Practice
Similar Empirical Results, Larger CIs

(a) Yoder (APSR, 2020)

(b) Bhavnani and Lee (AJPS, 2019)
Concluding Remarks

• Data sharing; data access
• DP protects individual privacy
• Enables inference to private database, not population
• Usually biased, no uncertainty estimates
• Fails to protect society from fallacious scientific conclusions

• Inferential validity
• A scientific statement: not necessarily correct, but must have:
  • known statistical properties
  &
  • valid uncertainty estimates

• Proposed algorithm
  • Generic: almost any statistical method or quantity of interest
  • Statistically unbiased, lower variance
  • Valid uncertainty estimates
  • Computationally efficient
• Solves political problems technologically

• Community based, Open Source Software: OpenDP.org
Concluding Remarks

• Data sharing \(\sim\) data access
Concluding Remarks

• Data sharing $\sim$ data access
  • DP protects individual privacy

The Algorithm in Practice
Concluding Remarks

• **Data sharing \(\sim\) data access**
  • DP protects individual privacy
  • Enables inference to private database, not population
Concluding Remarks

• **Data sharing $\leadsto$ data access**
  
  • DP protects individual privacy
  • Enables inference to private database, not population
  • Usually biased, no uncertainty estimates
Concluding Remarks

• **Data sharing ~ data access**
  - DP protects individual privacy
  - Enables inference to private database, not population
  - Usually biased, no uncertainty estimates
  - Fails to protect society from fallacious scientific conclusions

---

*The Algorithm in Practice*
Concluding Remarks

• Data sharing $\leadsto$ data access
  • DP protects individual privacy
  • Enables inference to private database, not population
  • Usually biased, no uncertainty estimates
  • Fails to protect society from fallacious scientific conclusions

• Inferential validity
Concluding Remarks

- **Data sharing → data access**
  - DP protects individual privacy
  - Enables inference to private database, not population
  - Usually biased, no uncertainty estimates
  - Fails to protect society from fallacious scientific conclusions

- **Inferential validity**
  - A scientific statement: not necessarily correct, but must have:
Concluding Remarks

• **Data sharing \(\sim\) data access**
  - DP protects individual privacy
  - Enables inference to private database, not population
  - Usually biased, no uncertainty estimates
  - Fails to protect society from fallacious scientific conclusions

• **Inferential validity**
  - A scientific statement: not necessarily correct, but must have:
  - known statistical properties & valid uncertainty estimates
Concluding Remarks

• Data sharing $\leadsto$ data access
  • DP protects individual privacy
  • Enables inference to private database, not population
  • Usually biased, no uncertainty estimates
  • Fails to protect society from fallacious scientific conclusions

• Inferential validity
  • A scientific statement: not necessarily correct, but must have:
    • known statistical properties & valid uncertainty estimates

• Proposed algorithm
Concluding Remarks

- **Data sharing → data access**
  - DP protects individual privacy
  - Enables inference to private database, not population
  - Usually biased, no uncertainty estimates
  - Fails to protect society from fallacious scientific conclusions

- **Inferential validity**
  - A scientific statement: not necessarily correct, but must have:
    - known statistical properties & valid uncertainty estimates

- **Proposed algorithm**
  - **Generic:** almost any statistical method or quantity of interest
Concluding Remarks

- Data sharing $\sim$ data access
  - DP protects individual privacy
  - Enables inference to private database, not population
  - Usually biased, no uncertainty estimates
  - Fails to protect society from fallacious scientific conclusions

- Inferential validity
  - A scientific statement: not necessarily correct, but must have:
    - known statistical properties & valid uncertainty estimates

- Proposed algorithm
  - Generic: almost any statistical method or quantity of interest
  - Statistically unbiased, lower variance
Concluding Remarks

• Data sharing ⇨ data access
  • DP protects individual privacy
  • Enables inference to private database, not population
  • Usually biased, no uncertainty estimates
  • Fails to protect society from fallacious scientific conclusions

• Inferential validity
  • A scientific statement: not necessarily correct, but must have:
    • known statistical properties & valid uncertainty estimates

• Proposed algorithm
  • Generic: almost any statistical method or quantity of interest
  • Statistically unbiased, lower variance
  • Valid uncertainty estimates
Concluding Remarks

- Data sharing $\rightsquigarrow$ data access
  - DP protects individual privacy
  - Enables inference to private database, not population
  - Usually biased, no uncertainty estimates
  - Fails to protect society from fallacious scientific conclusions

- Inferential validity
  - A scientific statement: not necessarily correct, but must have:
    - known statistical properties & valid uncertainty estimates

- Proposed algorithm
  - Generic: almost any statistical method or quantity of interest
  - Statistically unbiased, lower variance
  - Valid uncertainty estimates
  - Computationally efficient
Concluding Remarks

• Data sharing \(\sim\) data access
  • DP protects individual privacy
  • Enables inference to private database, not population
  • Usually biased, no uncertainty estimates
  • Fails to protect society from fallacious scientific conclusions

• Inferential validity
  • A scientific statement: not necessarily correct, but must have:
    • known statistical properties & valid uncertainty estimates

• Proposed algorithm
  • Generic: almost any statistical method or quantity of interest
  • Statistically unbiased, lower variance
  • Valid uncertainty estimates
  • Computationally efficient
  • Solves political problems technologically
Concluding Remarks

• **Data sharing** $\sim$ **data access**
  - DP protects individual privacy
  - Enables inference to private database, not population
  - Usually biased, no uncertainty estimates
  - Fails to protect society from fallacious scientific conclusions

• **Inferential validity**
  - A scientific statement: not necessarily correct, but must have:
    - known statistical properties & valid uncertainty estimates

• **Proposed algorithm**
  - **Generic**: almost any statistical method or quantity of interest
  - Statistically unbiased, lower variance
  - Valid uncertainty estimates
  - Computationally efficient
  - Solves political problems technologically

• **Community based, Open Source Software**: OpenDP.org
Articles, software, slides, videos: GaryKing.org/privacy
Articles, software, slides, videos: GaryKing.org/privacy

- Georgina Evans, Gary King, Margaret Schwenzfeier, and Abhradeep Thakurta. “Statistically Valid Inferences from Privacy Protected Data” *American Political Science Review*
• Georgina Evans, Gary King, Margaret Schwenzfeier, and Abhradeep Thakurta. “Statistically Valid Inferences from Privacy Protected Data” *American Political Science Review*

Articles, software, slides, videos: GaryKing.org/privacy

- Georgina Evans, Gary King, Margaret Schwenzfeier, and Abhradeep Thakurta. “Statistically Valid Inferences from Privacy Protected Data” *American Political Science Review*
- Georgina Evans, Gary King. “Statistically Valid Inferences from Differentially Private Data Releases, with Application to the Facebook URLs Dataset” *Political Analysis*
Appendix
Properties of Differential Privacy

- **Post-processing:** If $M(s, D)$ is DP, so is $f(M(s, D))$.

- **Useful for bias corrections**

- **Privacy risk quantified** ($\epsilon$), instead of 0/1 for re-ID.

- **Helpful mathematically; insufficient in applications**

- **Real privacy loss $\ll$ maximum privacy loss**

- **OK for worst case scenario; unhelpful in practice**

- **Privacy Budget**

  - **Composition:** $\epsilon_1$-DP and $\epsilon_2$-DP is $(\epsilon_1 + \epsilon_2)$-DP.

  - **Can limit maximum risks across analyses & researchers**

  - **When the budget is used, no new analyses can ever be run**

- **Without DP, we balance worries:**
  - **P-hacking; pre-registration (e.g., clinical trials, Mars lander)**
  - **Threats to inference; diagnostics, exploration, serendipity (e.g., observational data)**

- **With DP:** XXXXX

P-hacking, surveys treated like the Mars lander.
Properties of Differential Privacy

- **Post-processing:** if $M(s, D)$ is DP, so is $f[M(s, D)]$
Properties of Differential Privacy

• **Post-processing**: if \( M(s, D) \) is DP, so is \( f[M(s, D)] \)
  - Useful for bias corrections
Properties of Differential Privacy

- **Post-processing:** if $M(s, D)$ is DP, so is $f[M(s, D)]$
  - Useful for bias corrections
- **Privacy risk quantified** ($\epsilon$), instead of 0/1 for re-ID
Properties of Differential Privacy

- **Post-processing:** if $M(s, D)$ is DP, so is $f[M(s, D)]$
  - Useful for bias corrections
- **Privacy risk quantified** ($\epsilon$), instead of 0/1 for re-ID
  - Helpful mathematically; insufficient in applications
Properties of Differential Privacy

• Post-processing: if $M(s, D)$ is DP, so is $f[M(s, D)]$
  • Useful for bias corrections
• Privacy risk quantified ($\epsilon$), instead of 0/1 for re-ID
  • Helpful mathematically; insufficient in applications
• Real privacy loss $\ll$ maximum privacy loss
Properties of Differential Privacy

• **Post-processing**: if $M(s, D)$ is DP, so is $f[M(s, D)]$
  - Useful for bias corrections
• **Privacy risk quantified** ($\epsilon$), instead of 0/1 for re-ID
  - Helpful mathematically; insufficient in applications
• **Real privacy loss** $\ll$ maximum privacy loss
  - OK for worst case scenario; unhelpful in practice
Properties of Differential Privacy

• **Post-processing:** if $M(s, D)$ is DP, so is $f[M(s, D)]$
  • Useful for bias corrections
• **Privacy risk quantified** ($\epsilon$), instead of 0/1 for re-ID
  • Helpful mathematically; insufficient in applications
• **Real privacy loss** $\ll$ maximum privacy loss
  • OK for worst case scenario; unhelpful in practice
• **Privacy Budget**
Properties of Differential Privacy

• Post-processing: if $M(s, D)$ is DP, so is $f[M(s, D)]$
  • Useful for bias corrections
• Privacy risk quantified ($\epsilon$), instead of 0/1 for re-ID
  • Helpful mathematically; insufficient in applications
• Real privacy loss $\ll$ maximum privacy loss
  • OK for worst case scenario; unhelpful in practice
• Privacy Budget
  • Composition: $\epsilon_1$-DP and $\epsilon_2$-DP is $(\epsilon_1 + \epsilon_2)$-DP
Properties of Differential Privacy

• **Post-processing:** if $M(s, D)$ is DP, so is $f[M(s, D)]$
  • Useful for bias corrections
• **Privacy risk quantified** ($\epsilon$), instead of 0/1 for re-ID
  • Helpful mathematically; insufficient in applications
• **Real privacy loss** $\ll$ maximum privacy loss
  • OK for worst case scenario; unhelpful in practice
• **Privacy Budget**
  • **Composition:** $\epsilon_1$-DP and $\epsilon_2$-DP is $(\epsilon_1 + \epsilon_2)$-DP
  • Can limit maximum risks across analyses & researchers
Properties of Differential Privacy

- **Post-processing:** if $M(s, D)$ is DP, so is $f[M(s, D)]$
  - Useful for bias corrections
- **Privacy risk quantified** ($\epsilon$), instead of 0/1 for re-ID
  - Helpful mathematically; insufficient in applications
- **Real privacy loss** $\ll$ maximum privacy loss
  - OK for worst case scenario; unhelpful in practice
- **Privacy Budget**
  - **Composition:** $\epsilon_1$-DP and $\epsilon_2$-DP is $(\epsilon_1 + \epsilon_2)$-DP
  - **Can limit maximum risks** across analyses & researchers
  - When the budget is used, no new analyses can ever be run
Properties of Differential Privacy

• **Post-processing:** if $M(s, D)$ is DP, so is $f[M(s, D)]$
  • Useful for bias corrections
• **Privacy risk quantified** ($\epsilon$), instead of 0/1 for re-ID
  • Helpful mathematically; insufficient in applications
• **Real privacy loss** $\ll$ maximum privacy loss
  • OK for worst case scenario; unhelpful in practice
• **Privacy Budget**
  • **Composition:** $\epsilon_1$-DP and $\epsilon_2$-DP is $(\epsilon_1 + \epsilon_2)$-DP
  • Can limit maximum risks across analyses & researchers
  • When the budget is used, no new analyses can ever be run
• **Completely changes statistical best practices**
Properties of Differential Privacy

- **Post-processing**: if \( M(s, D) \) is DP, so is \( f[M(s, D)] \)
  - Useful for bias corrections
- **Privacy risk quantified** \((\epsilon)\), instead of 0/1 for re-ID
  - Helpful mathematically; insufficient in applications
- **Real privacy loss** \(\ll\) maximum privacy loss
  - OK for worst case scenario; unhelpful in practice
- **Privacy Budget**
  - **Composition**: \( \epsilon_1\)-DP and \( \epsilon_2\)-DP is \((\epsilon_1 + \epsilon_2)\)-DP
  - Can limit maximum risks across analyses & researchers
  - When the budget is used, *no new analyses can ever be run*
- **Completely changes statistical best practices**
  - Without DP, we balance worries:
Properties of Differential Privacy

- **Post-processing:** if \( M(s, D) \) is DP, so is \( f[M(s, D)] \)
  - Useful for bias corrections
- **Privacy risk quantified** \((\epsilon)\), instead of 0/1 for re-ID
  - Helpful mathematically; insufficient in applications
- **Real privacy loss** \(\ll\) maximum privacy loss
  - OK for worst case scenario; unhelpful in practice
- **Privacy Budget**
  - **Composition:** \( \epsilon_1\)-DP and \( \epsilon_2\)-DP is \( (\epsilon_1 + \epsilon_2)\)-DP
  - Can limit maximum risks across analyses & researchers
  - When the budget is used, no new analyses can ever be run
- **Completely changes statistical best practices**
  - **Without DP,** we balance worries:
    - P-hacking
Properties of Differential Privacy

- **Post-processing:** if $M(s, D)$ is DP, so is $f[M(s, D)]$
  - Useful for bias corrections
- **Privacy risk quantified** ($\epsilon$), instead of 0/1 for re-ID
  - Helpful mathematically; insufficient in applications
- **Real privacy loss $\ll$ maximum privacy loss**
  - OK for worst case scenario; unhelpful in practice
- **Privacy Budget**
  - **Composition:** $\epsilon_1$-DP and $\epsilon_2$-DP is $(\epsilon_1 + \epsilon_2)$-DP
  - Can limit maximum risks across analyses & researchers
  - When the budget is used, no new analyses can ever be run
- **Completely changes statistical best practices**
  - **Without DP,** we balance worries:
    - P-hacking
    - Threats to inference
Properties of Differential Privacy

• **Post-processing:** if $M(s, D)$ is DP, so is $f[M(s, D)]$
  - Useful for bias corrections

• **Privacy risk quantified** ($\epsilon$), instead of 0/1 for re-ID
  - Helpful mathematically; insufficient in applications

• **Real privacy loss** $\ll$ maximum privacy loss
  - OK for worst case scenario; unhelpful in practice

• **Privacy Budget**
  - **Composition:** $\epsilon_1$-DP and $\epsilon_2$-DP is $(\epsilon_1 + \epsilon_2)$-DP
  - Can limit maximum risks across analyses & researchers
  - When the budget is used, no new analyses can ever be run

• **Completely changes statistical best practices**
  - **Without DP,** we balance worries:
    - P-hacking $\rightsquigarrow$ pre-registration (e.g., clinical trials, Mars lander)
    - Threats to inference
Properties of Differential Privacy

- **Post-processing:** if $M(s, D)$ is DP, so is $f[M(s, D)]$
  - Useful for bias corrections
- **Privacy risk quantified** ($\epsilon$), instead of 0/1 for re-ID
  - Helpful mathematically; insufficient in applications
- **Real privacy loss $\ll$ maximum privacy loss**
  - OK for worst case scenario; unhelpful in practice
- **Privacy Budget**
  - **Composition:** $\epsilon_1$-DP and $\epsilon_2$-DP is $(\epsilon_1 + \epsilon_2)$-DP
  - Can limit maximum risks across analyses & researchers
  - When the budget is used, no new analyses can ever be run
- **Completely changes statistical best practices**
  - **Without DP,** we balance worries:
    - **P-hacking $\leadsto$** pre-registration (e.g., clinical trials, Mars lander)
    - **Threats to inference $\leadsto$** diagnostics, exploration, serendipity (e.g., observational data)
Properties of Differential Privacy

- **Post-processing:** if $M(s, D)$ is DP, so is $f[M(s, D)]$  
  - Useful for bias corrections
- **Privacy risk quantified ($\epsilon$), instead of 0/1 for re-ID**  
  - Helpful mathematically; insufficient in applications
- **Real privacy loss $\ll$ maximum privacy loss**  
  - OK for worst case scenario; unhelpful in practice
- **Privacy Budget**  
  - **Composition:** $\epsilon_1$-DP and $\epsilon_2$-DP is $(\epsilon_1 + \epsilon_2)$-DP  
  - Can limit maximum risks across analyses & researchers  
  - When the budget is used, no new analyses can ever be run
- **Completely changes statistical best practices**  
  - **Without DP,** we balance worries:  
    - P-hacking $\leadsto$ pre-registration (e.g., clinical trials, Mars lander)  
    - Threats to inference $\leadsto$ diagnostics, exploration, serendipity (e.g., observational data)
  - **With DP:**
Properties of Differential Privacy

• Post-processing: if $M(s, D)$ is DP, so is $f[M(s, D)]$
  • Useful for bias corrections
• Privacy risk quantified ($\epsilon$), instead of 0/1 for re-ID
  • Helpful mathematically; insufficient in applications
• Real privacy loss $\ll$ maximum privacy loss
  • OK for worst case scenario; unhelpful in practice
• Privacy Budget
  • Composition: $\epsilon_1$-DP and $\epsilon_2$-DP is $(\epsilon_1 + \epsilon_2)$-DP
  • Can limit maximum risks across analyses & researchers
  • When the budget is used, no new analyses can ever be run
• Completely changes statistical best practices
  • Without DP, we balance worries:
    • P-hacking $\leadsto$ pre-registration (e.g., clinical trials, Mars lander)
    • Threats to inference $\leadsto$ diagnostics, exploration, serendipity (e.g., observational data)
  • With DP: P-hacking,
Properties of Differential Privacy

• Post-processing: if $M(s, D)$ is DP, so is $f[M(s, D)]$
  • Useful for bias corrections
• Privacy risk quantified ($\epsilon$), instead of 0/1 for re-ID
  • Helpful mathematically; insufficient in applications
• Real privacy loss $\ll$ maximum privacy loss
  • OK for worst case scenerio; unhelpful in practice
• Privacy Budget
  • Composition: $\epsilon_1$-DP and $\epsilon_2$-DP is $(\epsilon_1 + \epsilon_2)$-DP
  • Can limit maximum risks across analyses & researchers
  • When the budget is used, no new analyses can ever be run
• Completely changes statistical best practices
  • Without DP, we balance worries:
    • P-hacking $\rightsquigarrow$ pre-registration (e.g., clinical trials, Mars lander)
    • Threats to inference $\rightsquigarrow$ diagnostics, exploration, serendipity (e.g., observational data)
  • With DP: P-hacking, surveys treated like the Mars lander