Statistically Valid Inferences from Privacy Protected Data

Gary King[1]

Institute for Quantitative Social Science
Harvard University

Quantitative Social Science Colloquium, Princeton University, 10/7/2022

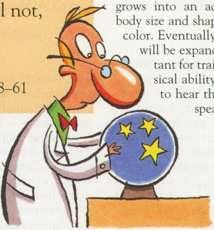
[1] GaryKing.org/privacy. Based on APSR/AJPS/PA articles with subsets of {Georgie Evans, Meg Schwenzfeier, Abhradeep Thakurta, Adam D. Smith}

Science Magazine, 1995

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 knockout 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 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

ILLUSTRATIONS BY TERRY E. SMITH

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

Progress in Social Science

- What did 60 scientists forecast in 1995?

Progress in Social Science

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

Progress in Social Science

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

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.

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

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

Progress in Social Science

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

Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via "constitutional design")

Convincing Facebook to Make Data Available

Solving a Political Problem Technologically (via "constitutional design")

- Gary visits Facebook to persuade them to make data available

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.

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:

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)

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 $\rightsquigarrow$ 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 $\rightsquigarrow$ 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 $\rightsquigarrow$ 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 $\rightsquigarrow$ Data Access Regime

Data Sharing Regime $\rightsquigarrow$ Data Access Regime

Solving Another Political Problem Technologically (via CS & Statistics)

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)

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)

- 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
- 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

- 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!
- 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
- 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,
- 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 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,
- 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

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 $\rightsquigarrow$ 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,

- 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, restricted viewing, paired programmer models, etc.
- 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, restricted viewing, paired programmer models, etc.
  - Trusting researchers fails spectacularly at times (C.A.!)
- 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, 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 $\rightsquigarrow$ 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;

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 $\rightsquigarrow$ 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,

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

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:

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

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 $\rightsquigarrow$ 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
  - $\approx$ differential 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
  - $\approx$ differential privacy (seems to satisfy regulators et al.)

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
  - $\approx$ differential privacy (seems to satisfy regulators et al.)
  - New Problem:

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
  - $\approx$ differential privacy (seems to satisfy regulators et al.)
  - New Problem: Most DP algorithms are statistically invalid!

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
  - $\approx$ differential privacy (seems to satisfy regulators et al.)
  - New Problem: Most DP algorithms are statistically invalid!
    - *unknown* statistical properties (usually *biased*)

Data Sharing Regime $\rightsquigarrow$ 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
  - $\approx$ 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

Theories of Inference: Statistics vs. CS

Population
⋮
Rocío
John
Marc
Brandon
Yu Xie
Gleason
Saad
Leonard
Kristopher
Zhou

Mean income: \$48

Quantity of Interest

Theories of Inference: Statistics vs. CS

Population	Sample
⋮	✗
Rocío	✓
John	✓
Marc	✓
Brandon	✓
Yu Xie	✓
Gleason	✓
Saad	✓
Leonard	✓
Kristopher	✓
Zhou	✓
Mean income: $48	

Quantity of Interest

Theories of Inference: Statistics vs. CS

Population	Sample	$
⋮	✗	?
Rocío	✓	122
John	✓	76
Marc	✓	145
Brandon	✓	96
Yu Xie	✓	86
Gleason	✓	127
Saad	✓	72
Leonard	✓	132
Kristopher	✓	95
Zhou	✓	134
Mean income: $48		$108

Classical Inference

Quantity of Interest

Usually no direct relevance

Theories of Inference: Statistics vs. CS

Population	Sample	$
⋮	✗	?
Rocío	✓	122
John	✓	76
Marc	✓	145
Brandon	✓	96
Yu Xie	✓	86
Gleason	✓	127
Saad	✓	72
Leonard	✓	132
Kristopher	✓	95
Zhou	✓	134
$48		$108

Mean income:

Classical Inference

Quantity of Interest

Usually no direct relevance

Theories of Inference: Statistics vs. CS

Population	Sample	$	+Privacy
⋮	✗	?	
Rocío	✓	122	Noise & Censoring
John	✓	76	
Marc	✓	145	
Brandon	✓	96	
Yu Xie	✓	86	
Gleason	✓	127	
Saad	✓	72	
Leonard	✓	132	
Kristopher	✓	95	
Zhou	✓	134	
Mean income: $48		$108	

Classical Inference

Quantity of Interest

Usually no direct relevance

Theories of Inference: Statistics vs. CS

Population	Sample	$	+Privacy	=dp$
⋮	✗	?		
Rocío	✓	122		85
John	✓	76		103
Marc	✓	145		75
Brandon	✓	96		113
Yu Xie	✓	86	Noise & Censoring	125
Gleason	✓	127		97
Saad	✓	72		101
Leonard	✓	132		128
Kristopher	✓	95		83
Zhou	✓	134		201
Mean income: $48		$108		$111

$48 $\xleftarrow{\text{Classical Inference}}$ $108 $\xleftarrow{\text{Query-Response}}$ $111

Quantity of Interest

Usually no direct relevance

No direct relevance

Theories of Inference: Statistics vs. CS

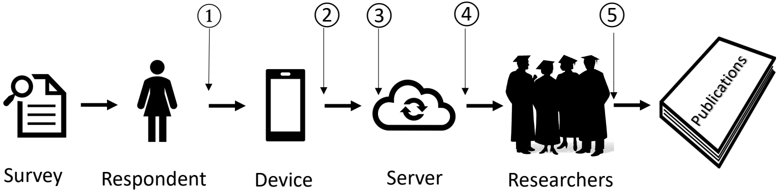
Population	Sample	$	+Privacy	=dp$
⋮	✗	?		
Rocío	✓	122		85
John	✓	76		103
Marc	✓	145		75
Brandon	✓	96		113
Yu Xie	✓	86	Noise & Censoring	125
Gleason	✓	127		97
Saad	✓	72		101
Leonard	✓	132		128
Kristopher	✓	95		83
Zhou	✓	134		201
Mean income: $48		$108		$111

Classical Inference

Query-Response

Statistically Valid Inferences from Privacy Protected Data

Protecting Survey Data



Differential Privacy and its Inferential Challenges

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

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 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
  - $M(\text{mean}, D) = \frac{1}{n}\sum_{i=1}^{n} c(y_i, \Lambda) + N\left(0, \frac{8\Lambda}{n\epsilon}\right)$ $\qquad (\Lambda, n, \epsilon$ 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
  - $M(\text{mean}, D) = \frac{1}{n} \sum_{i=1}^{n} c(y_i, \Lambda) + N\left(0, \frac{8\Lambda}{n\epsilon}\right)$ $(\Lambda, n, \epsilon$ known$)$
  - Or: mess with gradients, $X_i' X_i$, 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
  - $M(\text{mean}, D) = \frac{1}{n} \sum_{i=1}^{n} c(y_i, \Lambda) + N\left(0, \frac{8\Lambda}{n\epsilon}\right)$ $\qquad$ (Λ, n, ϵ 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

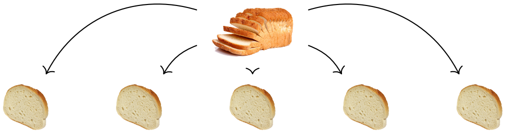
The Algorithm in Practice

A Differentially Private Estimator

A Differentially Private Estimator

Private data

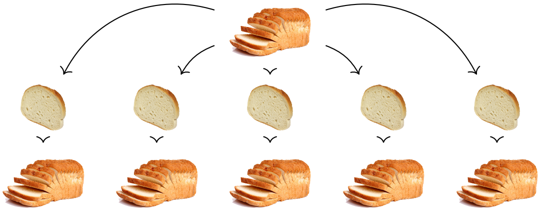
A Differentially Private Estimator



Private data

Partition

A Differentially Private Estimator

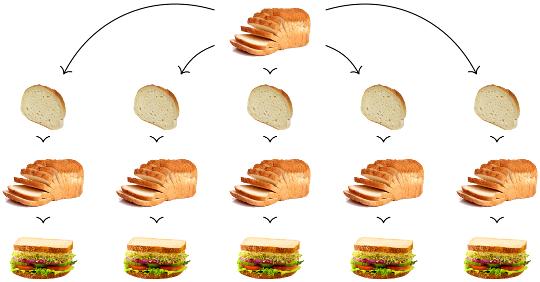


Private data

Partition

Bag of little bootstraps

A Differentially Private Estimator



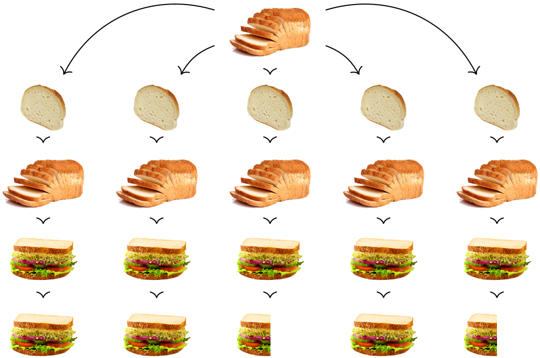
Private data

Partition

Bag of little bootstraps

Estimator

A Differentially Private Estimator



Private data

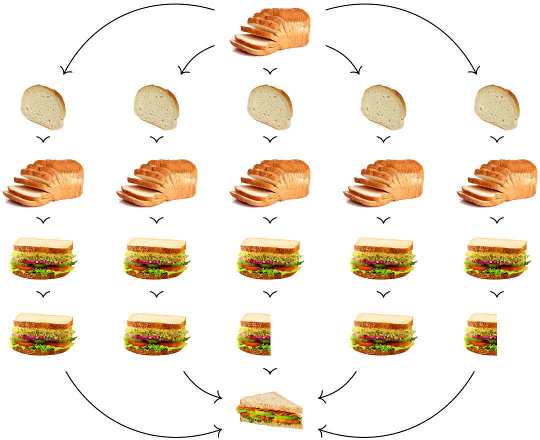
Partition

Bag of little bootstraps

Estimator

Censor

A Differentially Private Estimator



Private data

Partition

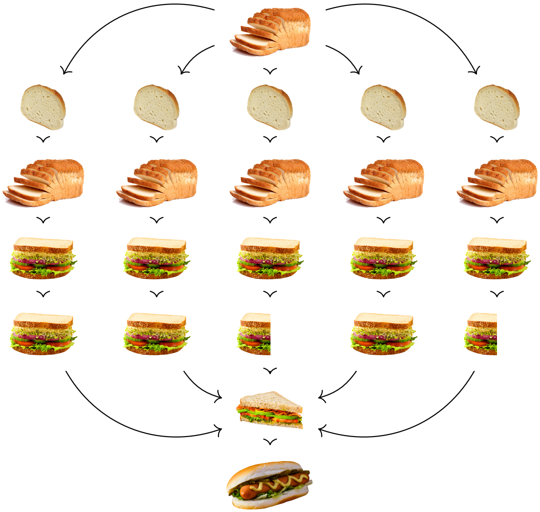
Bag of little bootstraps

Estimator

Censor

Average

A Differentially Private Estimator



Private data

Partition

Bag of little bootstraps

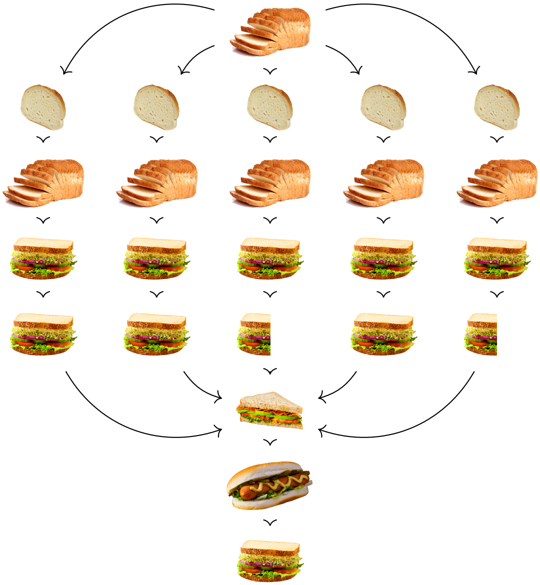
Estimator

Censor

Average

Noise

A Differentially Private Estimator



Private data

Partition

Bag of little bootstraps

Estimator

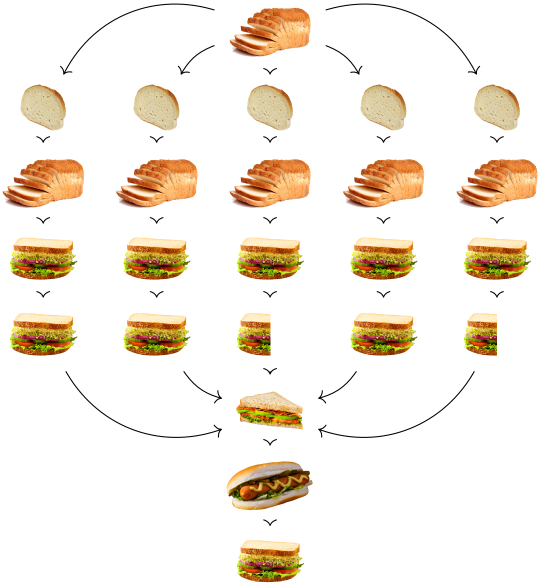
Censor

Average

Noise

Bias Correction

A Differentially Private Estimator



Private data

Partition

Bag of little bootstraps

Estimator

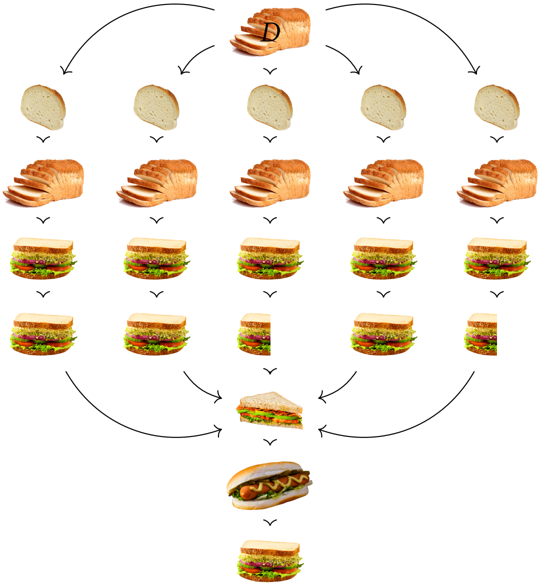
Censor

Average

Noise

Bias Correction
(& variance estimation)

A Differentially Private Estimator



Private data

Partition

Bag of little bootstraps

Estimator

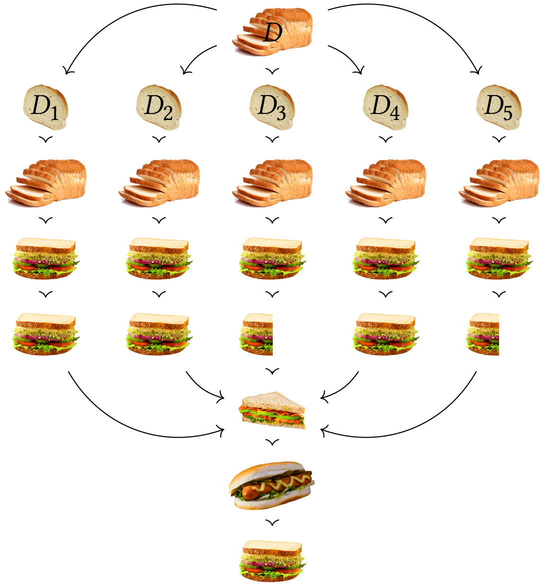
Censor

Average

Noise

Bias Correction
(& variance estimation)

A Differentially Private Estimator



Private data

Partition

Bag of little bootstraps

Estimator

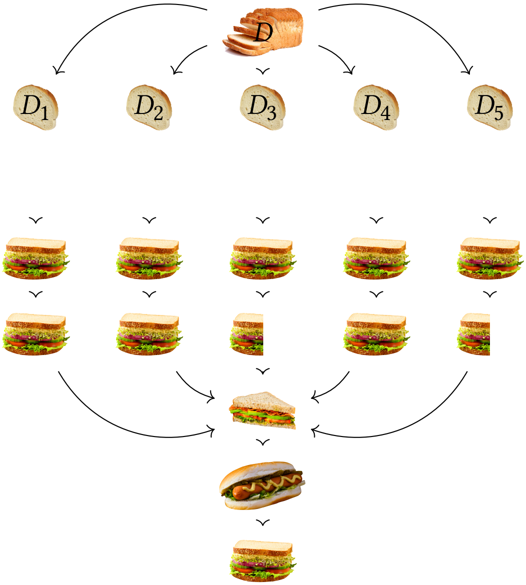
Censor

Average

Noise

Bias Correction
(& variance estimation)

A Differentially Private Estimator



Private data

Partition

Bag of little bootstraps

Estimator

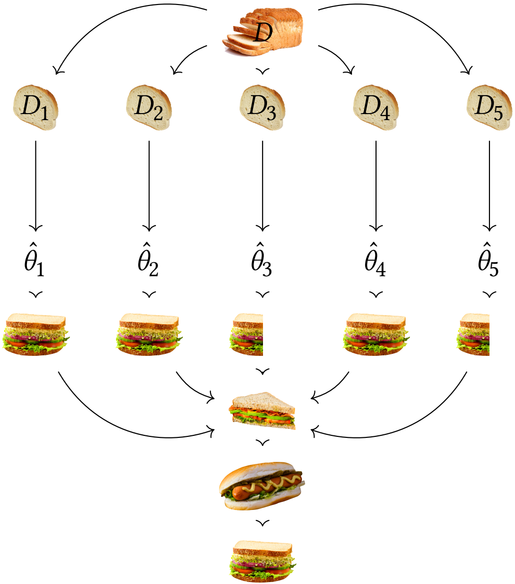
Censor

Average

Noise

Bias Correction
(& variance estimation)

A Differentially Private Estimator



Private data

Partition

Bag of little bootstraps

Estimator

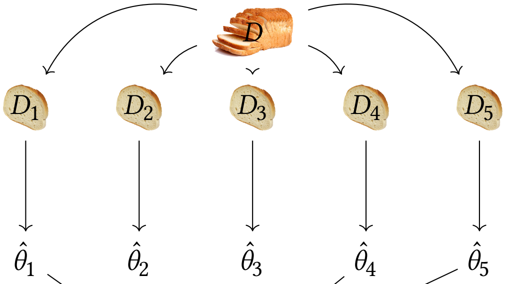
Censor

Average

Noise

Bias Correction
(& variance estimation)

A Differentially Private Estimator



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

Private data

Partition

Bag of little bootstraps

Estimator

Censor

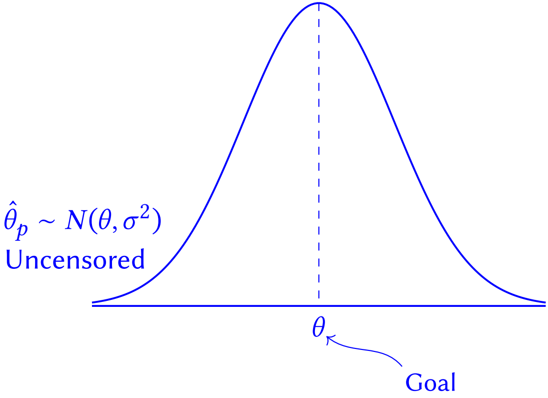
Average

Noise

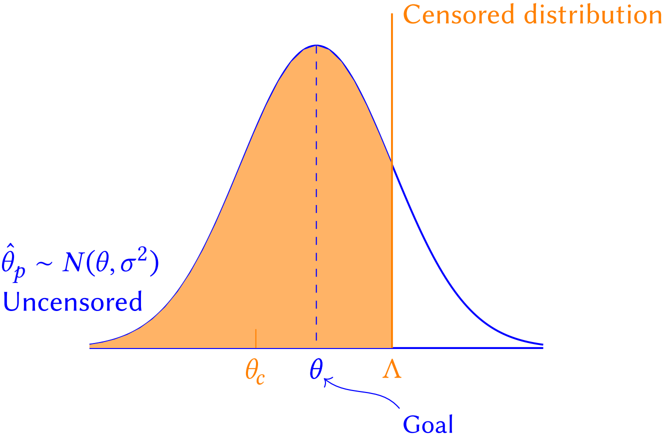
Bias Correction
(& variance estimation)

Bias Correction of: $\hat{\theta}^{\mathrm{dp}} = \frac{1}{P}\sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N\left(0, \frac{8\Lambda}{P\epsilon}\right)$ (Λ, P, ϵ known)

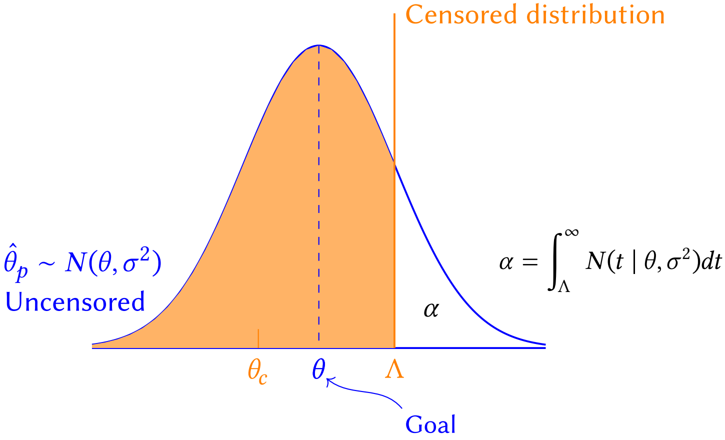
Bias Correction of: $\hat{\theta}^{\mathrm{dp}} = \frac{1}{P}\sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N\left(0, \frac{8\Lambda}{P\epsilon}\right)$ (Λ, P, ϵ known)



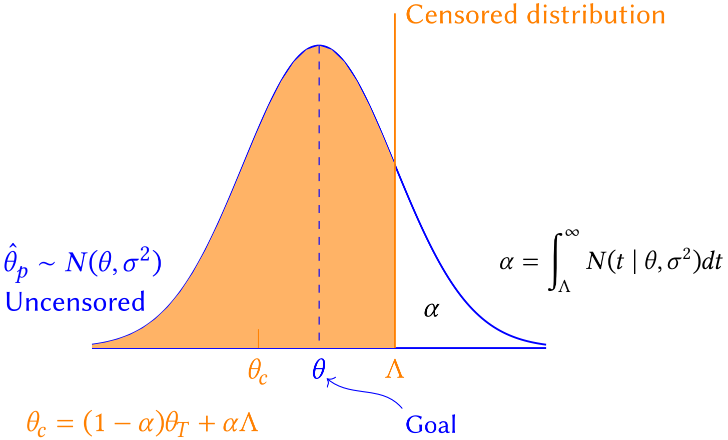
Bias Correction of: $\hat{\theta}^{\text{dp}} = \frac{1}{P}\sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N\left(0, \frac{8\Lambda}{P\epsilon}\right)$ (Λ, P, ϵ known)



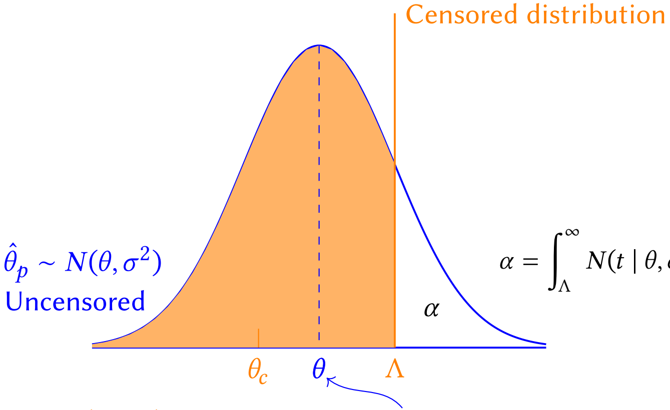
Bias Correction of: $\hat{\theta}^{\mathrm{dp}} = \frac{1}{P}\sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N\left(0, \frac{8\Lambda}{P\epsilon}\right)$ (Λ, P, ϵ known)



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



Bias Correction of: $\hat{\theta}^{\mathrm{dp}} = \frac{1}{P}\sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N\left(0, \frac{8\Lambda}{P\epsilon}\right)$ (Λ, P, ϵ known)

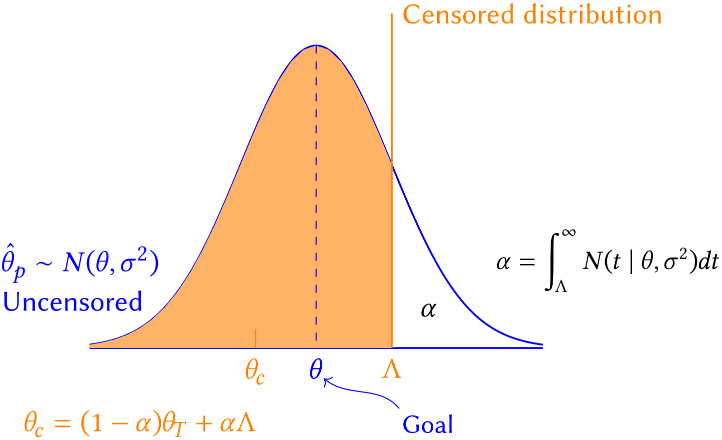


$$\alpha = \int_{\Lambda}^{\infty} N(t \mid \theta, \sigma^2)\,dt$$

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

Equations: 2

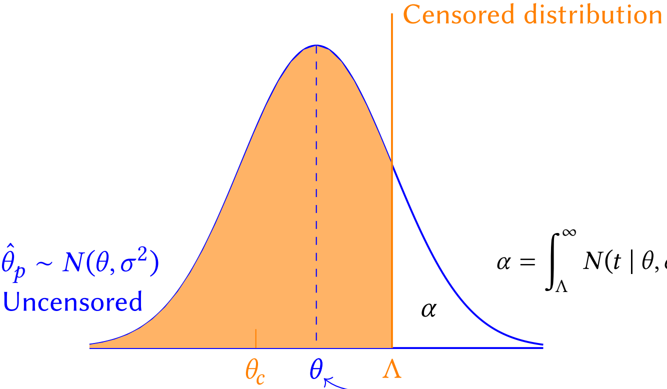
Bias Correction of: $\hat{\theta}^{\mathrm{dp}} = \frac{1}{P}\sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N\left(0, \frac{8\Lambda}{P\epsilon}\right)$ (Λ, P, ϵ known)



Equations: 2

Unknowns: θ, σ^2, α, θ_c

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



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

Goal

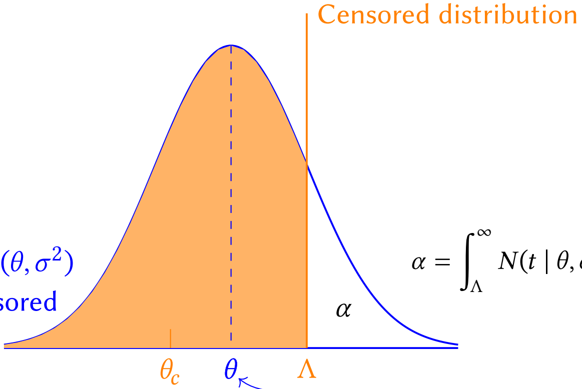
$\theta_c = (1-\alpha)\theta_T + \alpha\Lambda$

Disclose: $\hat{\theta}^{\text{dp}}$

Equations: 2

Unknowns: θ, σ^2, α, $\cancel{\theta_c}$

Bias Correction of: $\hat{\theta}^{\mathrm{dp}} = \frac{1}{P}\sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda) + N\left(0, \frac{8\Lambda}{P\epsilon}\right)$ (Λ, P, ϵ known)



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

Uncensored

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

Goal

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

Disclose: $\hat{\theta}^{\mathrm{dp}}$, $\hat{\alpha}^{\mathrm{dp}}$

Equations: 2

Unknowns: θ, σ^2, $\cancel{\alpha}$, $\cancel{\theta_c}$

Variance Estimation

Variance Estimation

- Simulate estimates via standard (Clarify) procedures:

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

Variance Estimation

- Simulate estimates via standard (Clarify) procedures:

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

Functions of disclosed params

Variance Estimation

- Simulate estimates via standard (Clarify) procedures:

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

Functions of disclosed params

- Bias correct simulated params:

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

Variance Estimation

- Simulate estimates via standard (Clarify) procedures:

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

Functions of disclosed params

- Bias correct simulated params:

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

- Standard error: Standard deviation of $\tilde{\theta}^{\text{dp}}$ over simulations

Variance Estimation

- Simulate estimates via standard (Clarify) procedures:

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

Functions of disclosed params

- Bias correct simulated params:

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

- Standard error: Standard deviation of $\tilde{\theta}^{\text{dp}}$ over simulations
- Bias correction: reduces bias *and* variance

Solving Political Problems Technologically

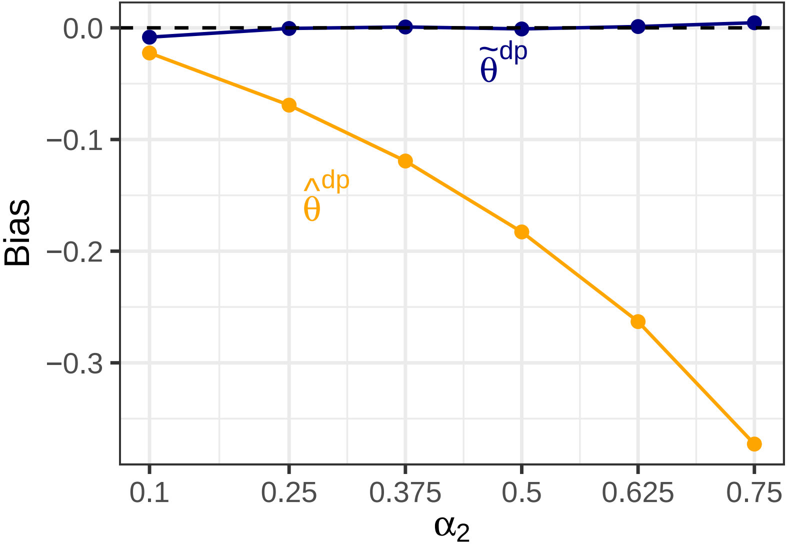
Differential Privacy & Inferential Validity

A General Purpose, Statistically Valid DP Algorithm

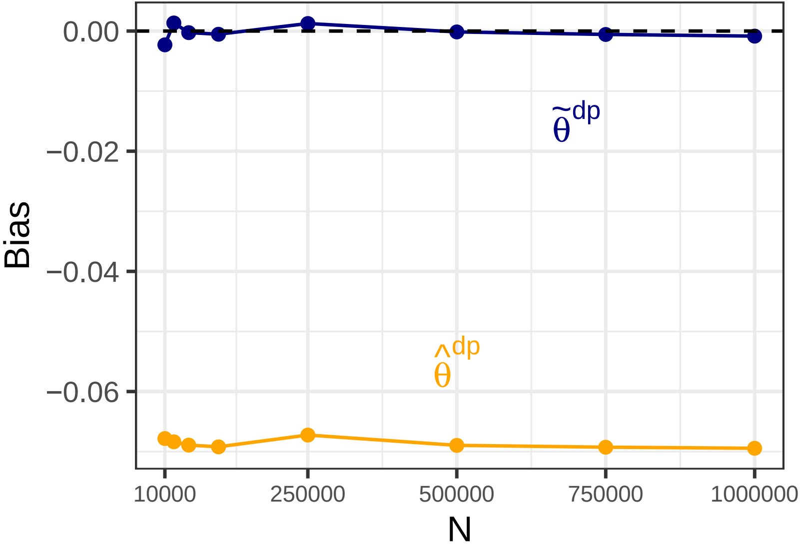
The Algorithm in Practice

Simulations: Finite Sample Evaluation

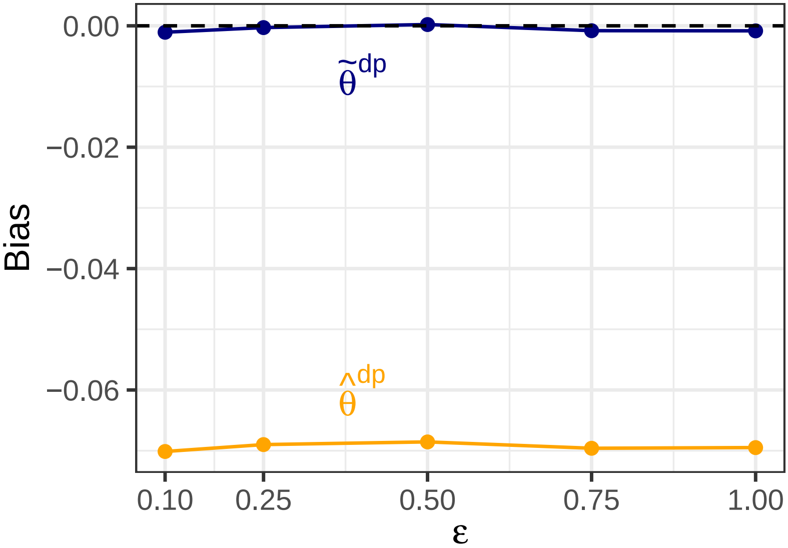
Simulations: Finite Sample Evaluation



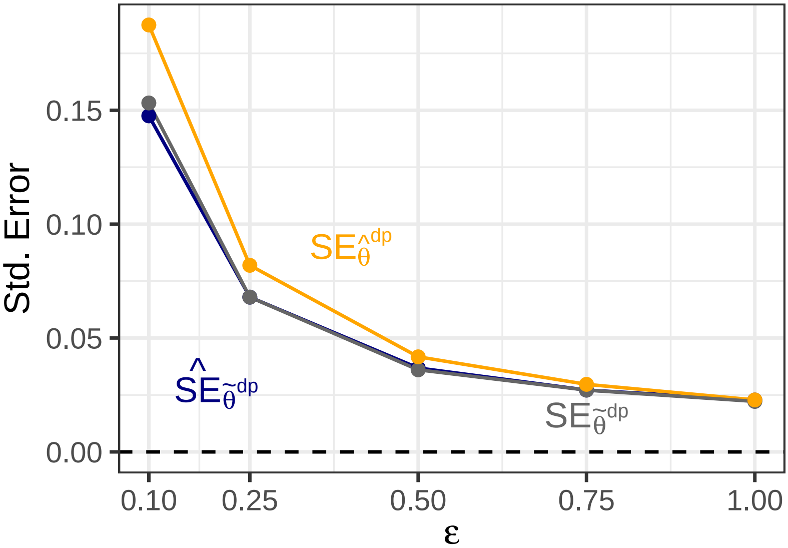
Simulations: Finite Sample Evaluation



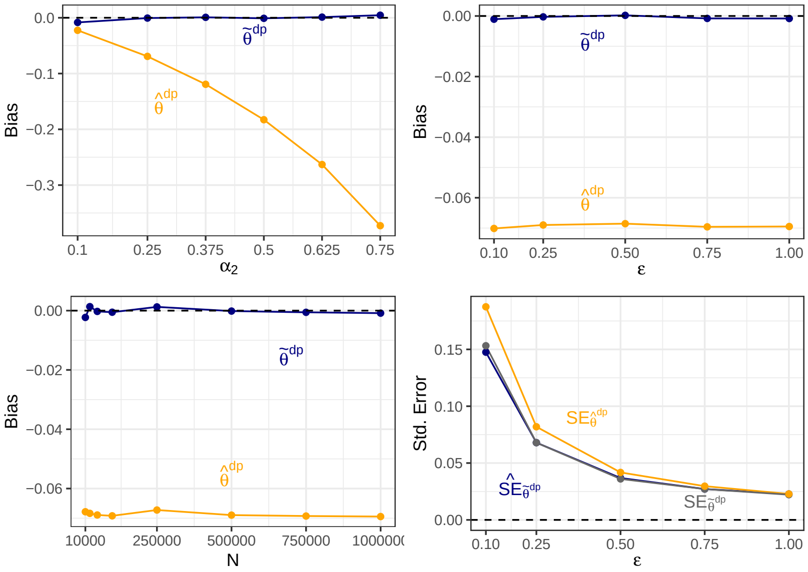
Simulations: Finite Sample Evaluation



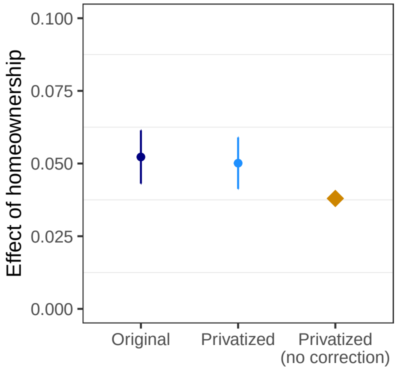
Simulations: Finite Sample Evaluation



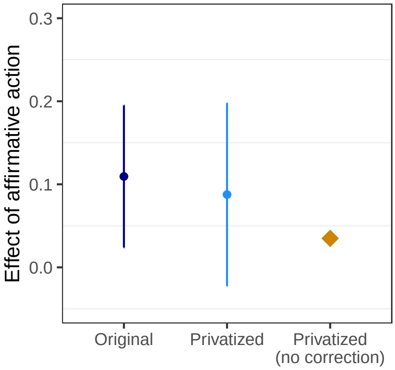
Simulations: Finite Sample Evaluation



Similar Empirical Results, Larger CIs



(a) Yoder (APSR, 2020)



(b) Bhavnani and Lee (AJPS, 2019)

Concluding Remarks

Concluding Remarks

- Data sharing $\rightsquigarrow$ data access

Concluding Remarks

- Data sharing $\rightsquigarrow$ data access
  - DP protects individual privacy

Concluding Remarks

- Data sharing $\rightsquigarrow$ data access
  - DP protects individual privacy
  - Enables inference to private database, not population

Concluding Remarks

- Data sharing $\rightsquigarrow$ data access
  - DP protects individual privacy
  - Enables inference to private database, not population
  - Usually biased, no 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

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

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:

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

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

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

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

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

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 $\rightsquigarrow$ 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
  - Solves political problems technologically

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
  - 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*

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, Adam D. Smith, Abhradeep Thakurta. "Differentially Private Survey Research" *American Journal of Political Science*

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, Adam D. Smith, Abhradeep Thakurta. "Differentially Private Survey Research" *American Journal of Political Science*
- 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

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 (ϵ), 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 (ϵ), 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 (ϵ), 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 (ϵ), 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

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 (ϵ), 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

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 (ϵ), 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: ϵ_1-DP and ϵ_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 (ϵ), 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: ϵ_1-DP and ϵ_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 (ϵ), 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: ϵ_1-DP and ϵ_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 (ϵ), 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: ϵ_1-DP and ϵ_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 (ϵ), 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: ϵ_1-DP and ϵ_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 (ϵ), 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: ϵ_1-DP and ϵ_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 (ϵ), 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: ϵ_1-DP and ϵ_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 (ϵ), 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: ϵ_1-DP and ϵ_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 (ϵ), 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: ϵ_1-DP and ϵ_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)

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 (ϵ), 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: ϵ_1-DP and ϵ_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:

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 (ϵ), 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: ϵ_1-DP and ϵ_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~~,

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 (ϵ), 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: ϵ_1-DP and ϵ_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