



DO NO HARM GUIDE

APPLYING EQUITY AWARENESS IN DATA PRIVACY METHODS

BY CLAIRE MCKAY BOWEN AND JOSHUA SNOKE

Joshua Snoke, PhD

Statistician | RAND Corporation

 jsnoke@rand.org



Talk Overview

- **Motivation and Background**

What inspired this guide?

How did we approach the study?

- **Defining Equity**

How can we define equity in the context of privacy?

- **Where do we go from here?**

An aspirational proposal and concluding recommendations

DO NO HARM GUIDE

APPLYING EQUITY AWARENESS
IN DATA PRIVACY METHODS

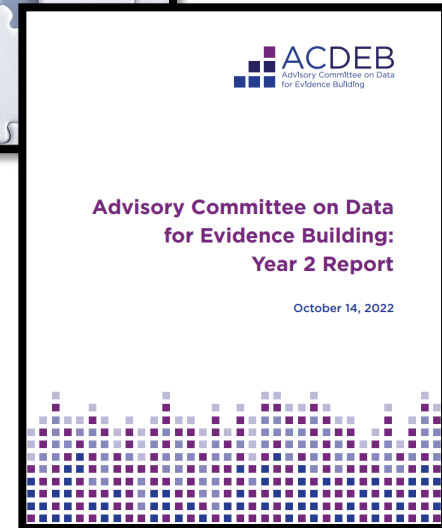
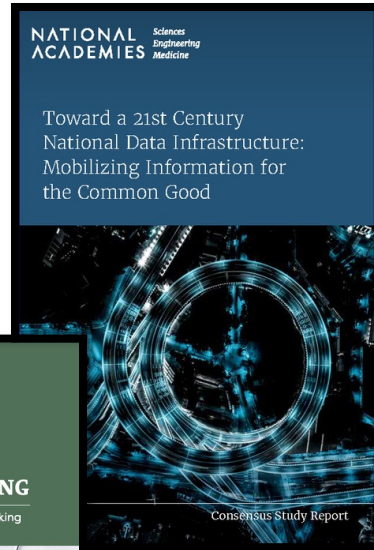
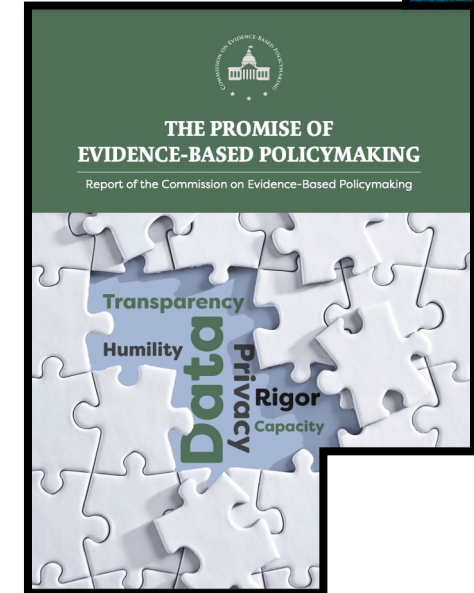
BY CLAIRE MCKAY BOWEN AND JOSHUA SNOKE

MARCH 2023

Why write this guide?

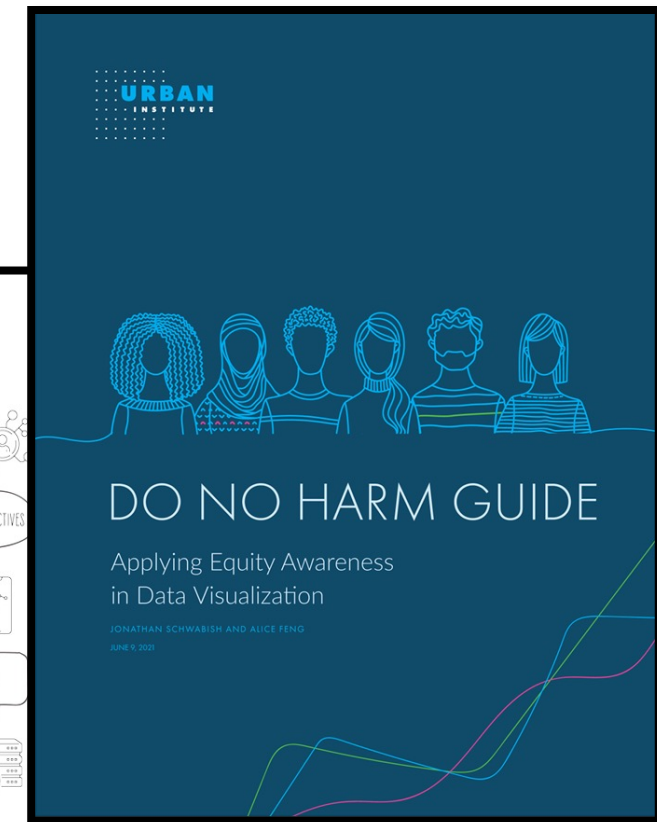
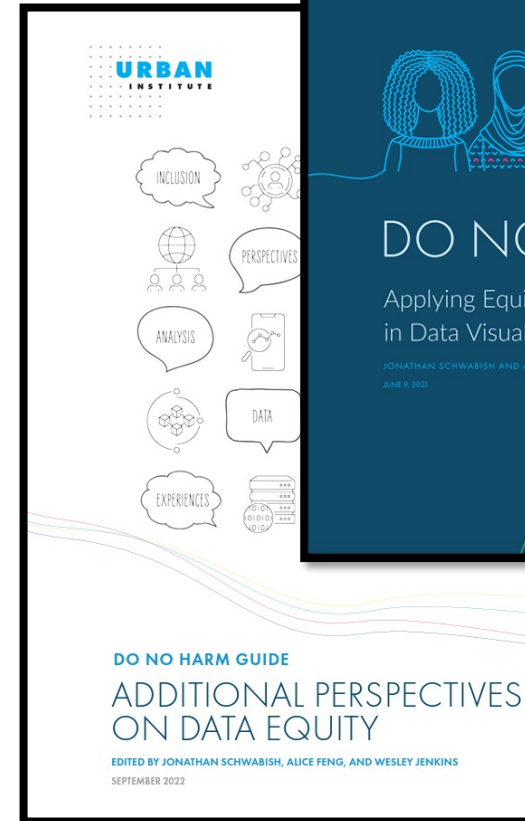
Motivation

- Lots conversations about equity *and* privacy (e.g., Biden's Executive Orders, CNSTAT reports, Evidence Act)
- Fewer conversations about equity *in* privacy
- Caveats about our guide:
 - High level work, primarily qualitative
 - Exploratory and foundation laying
 - Recommendations are aspirational



Background

- Volume 4 of the Tableau funded *Do No Harm Guide* series
- Target audience are those with some familiarity with statistical data privacy (SDP)
- Guide includes:
 - Literature review (mostly technical)
 - 9 subject matter expert interviews
 - 3 illustrative examples, showing the impacts of various SDP methods



“

We have norms on equity in decisionmaking. We have norms on equity in privacy loss. However, we're seeing a conflict between these two norms. We cannot have both equity in the decisionmaking process and privacy.

ASHWIN MACHANAVAJJHALA, ASSOCIATE PROFESSOR AT DUKE UNIVERSITY AND COFOUNDER OF TUMULT LABS

DO NO HARM GUIDE: APPLYING EQUITY AWARENESS IN DATA PRIVACY METHODS





JÖRG DRECHSLER

Distinguished researcher, Department for Statistical Methods, Institute for Employment Research, and associate research professor, Joint Program in Survey Methodology, University of Maryland



SAKI KINNEY

Senior research statistician, RTI International



TOM KRENZKE

Vice president in statistics and data science, Westat



ASHWIN MACHANAVAJJHALA

Founder and chief scientist, Tumult Labs, and associate professor, Department of Computer Science, Duke University



MAURICIO ORTIZ

Associate director for regional economics, Bureau of Economic Analysis



ROLANDO A. RODRÍGUEZ

Chief, Disclosure Avoidance Research and Applications Group, US Census Bureau



ALEKSANDRA SLAVKOVIĆ

Professor of statistics and associate dean for graduate education, Eberly College of Science at Penn State



ADAM SMITH

Professor of computer science and engineering, Boston University



DANIELL TOTH

Senior research mathematical statistician, US Bureau of Labor Statistics



JÖRG DRECHSLER

Distinguished researcher, Department for Statistical Methods, Institute for Employment Research, and associate research professor, Joint Program in Survey Methodology, University of Maryland



SAKI KINNEY

Senior research statistician, RTI International



TOM KRENZKE

Vice president in statistics and data science, Westat



ASHWIN MACHANAVAJJHALA

Founder and chief scientist, Tumult Labs, and associate professor, Department of Computer Science, Duke University



MAURICIO ORTIZ

Associate director for regional economics, Bureau of Economic Analysis



ROLANDO A. RODRÍGUEZ

Chief, Disclosure Avoidance Research and Applications Group, US Census Bureau



ALEKSANDRA SLAVKOVIĆ

Professor of statistics and associate dean for graduate education, Eberly College of Science at Penn State



ADAM SMITH

Professor of computer science and engineering, Boston University



DANIELL TOTH

Senior research mathematical statistician, US Bureau of Labor Statistics

THANK YOU!

How do we define equity?

“



What does equal utility mean and what does equal privacy loss mean? **There are multiple ways of defining equal.**

ALEKSANDRA SLAVKOVIĆ, PROFESSOR OF STATISTICS AND ASSOCIATE DEAN FOR GRADUATE EDUCATION IN EBERLY COLLEGE OF SCIENCE AT PENN STATE
DO NO HARM GUIDE: APPLYING EQUITY AWARENESS IN DATA PRIVACY METHODS



What are the equity concepts within SDP?

1. Disparities in the privacy loss and utility
2. Definitions and preferences of privacy loss and utility
3. Defining groups and representation
4. Resource allocation
5. Related processes
6. Misalignment with other efforts
7. Equity in the research field

1. Disparities in the Privacy Loss and Utility

Conceptual Separate Group Curves Showing Privacy Loss–Utility Tradeoffs

A fictitious dataset with privacy loss and utility values from 0 to 1 that are unitless

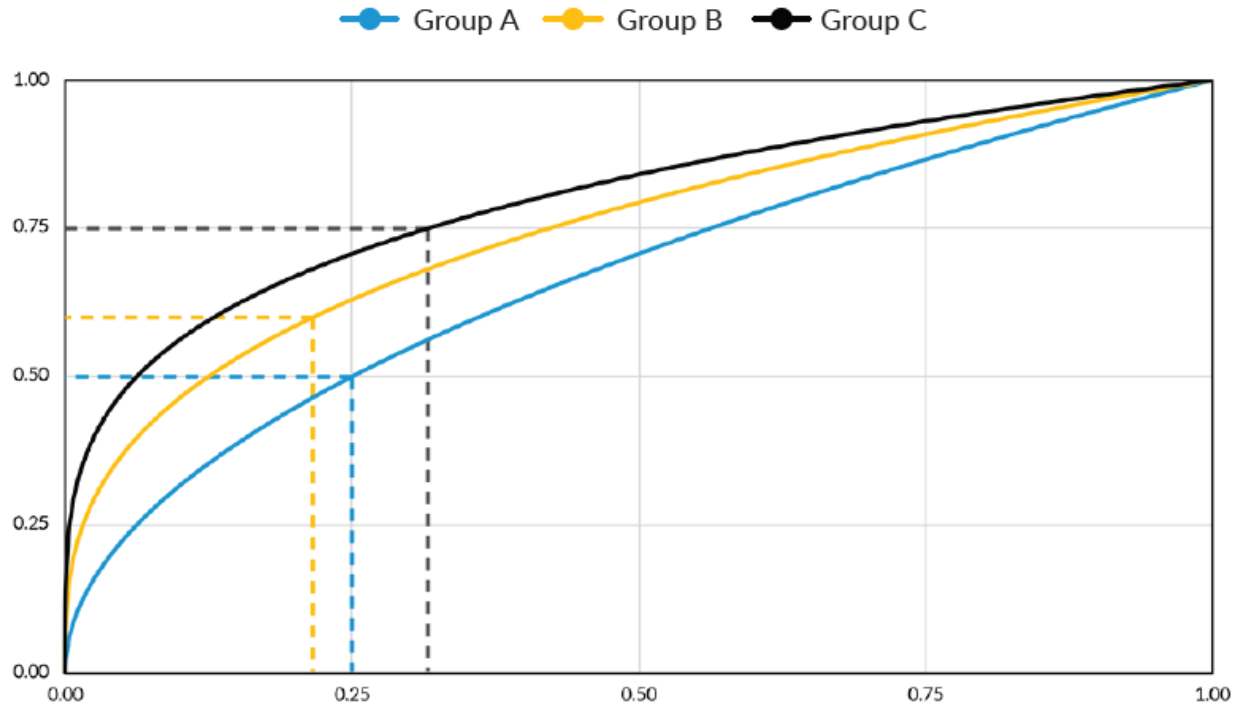


Figure 2a demonstrates that different groups can have different privacy loss and utility curves.

Source: Authors' illustration using hypothetical data.

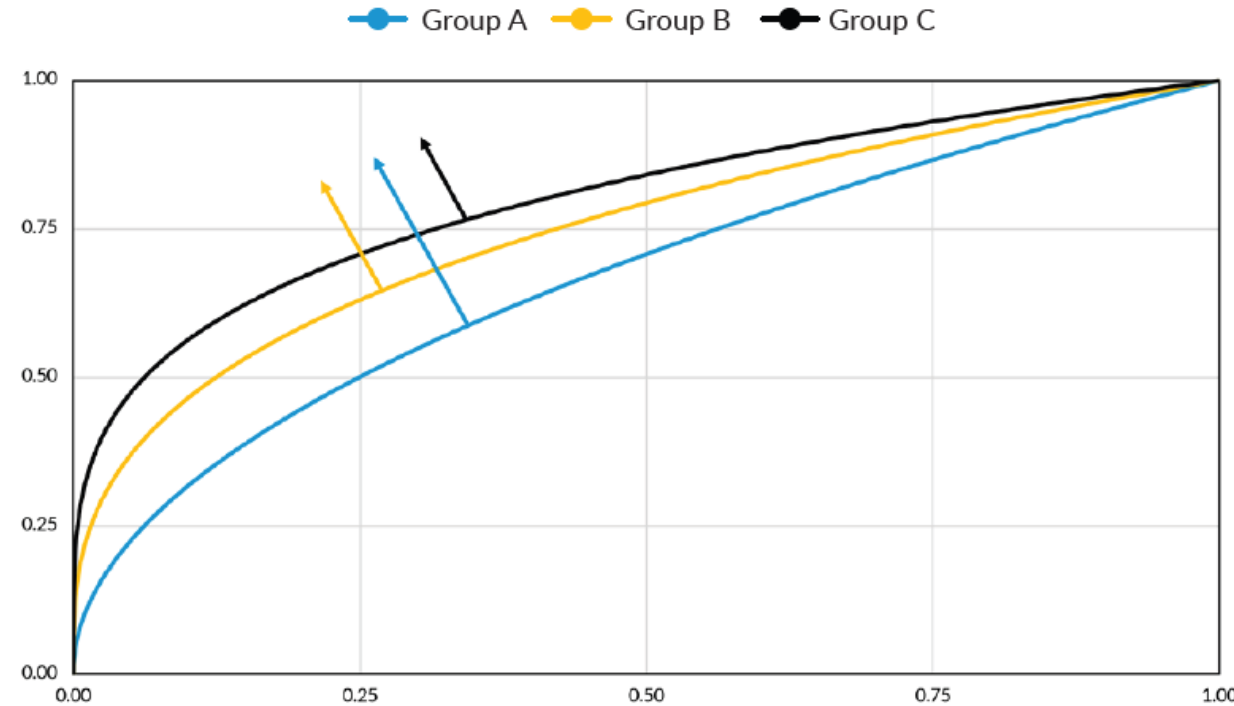


Figure 2b shows the goal of pushing all curves upward toward the top-left corner for the best privacy loss and utility tradeoff.

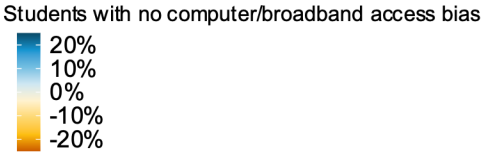
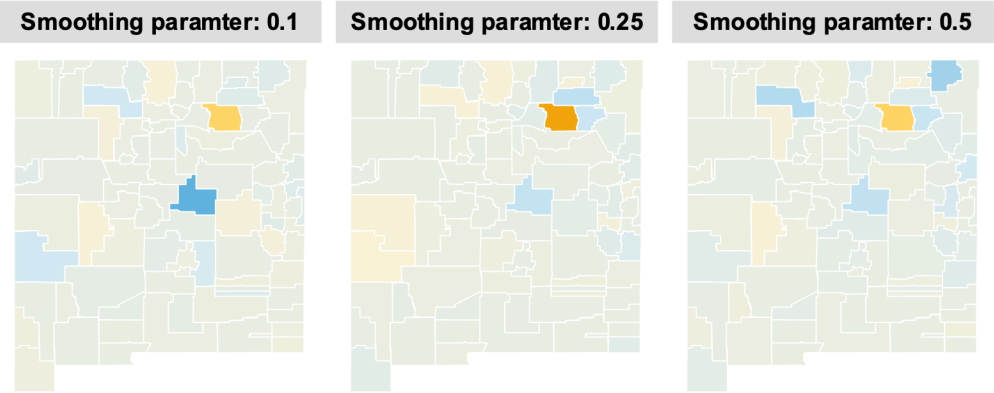
A brief aside on technical solutions

Technical Solutions Alone Will Not Suffice

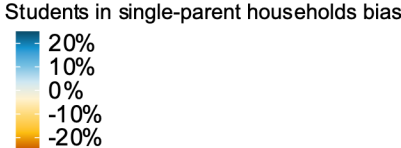
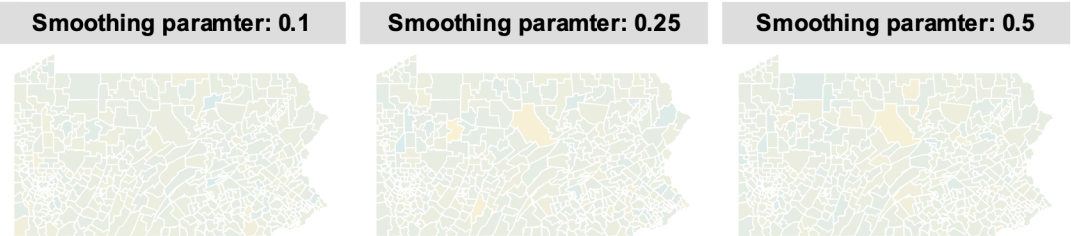
- No method is inherently equitable
- Smaller and less homogeneous groups are fundamentally harder to protect
- Simple examples to highlight these constraints
 - Apply suppression, synthesis, and differential privacy to ACS data for New Mexico and Pennsylvania school districts
 - Replicate analyses from Blagg et al. (2020) “Mapping Student Needs During COVID-19.”

Differing Tradeoffs for Synthetic Data

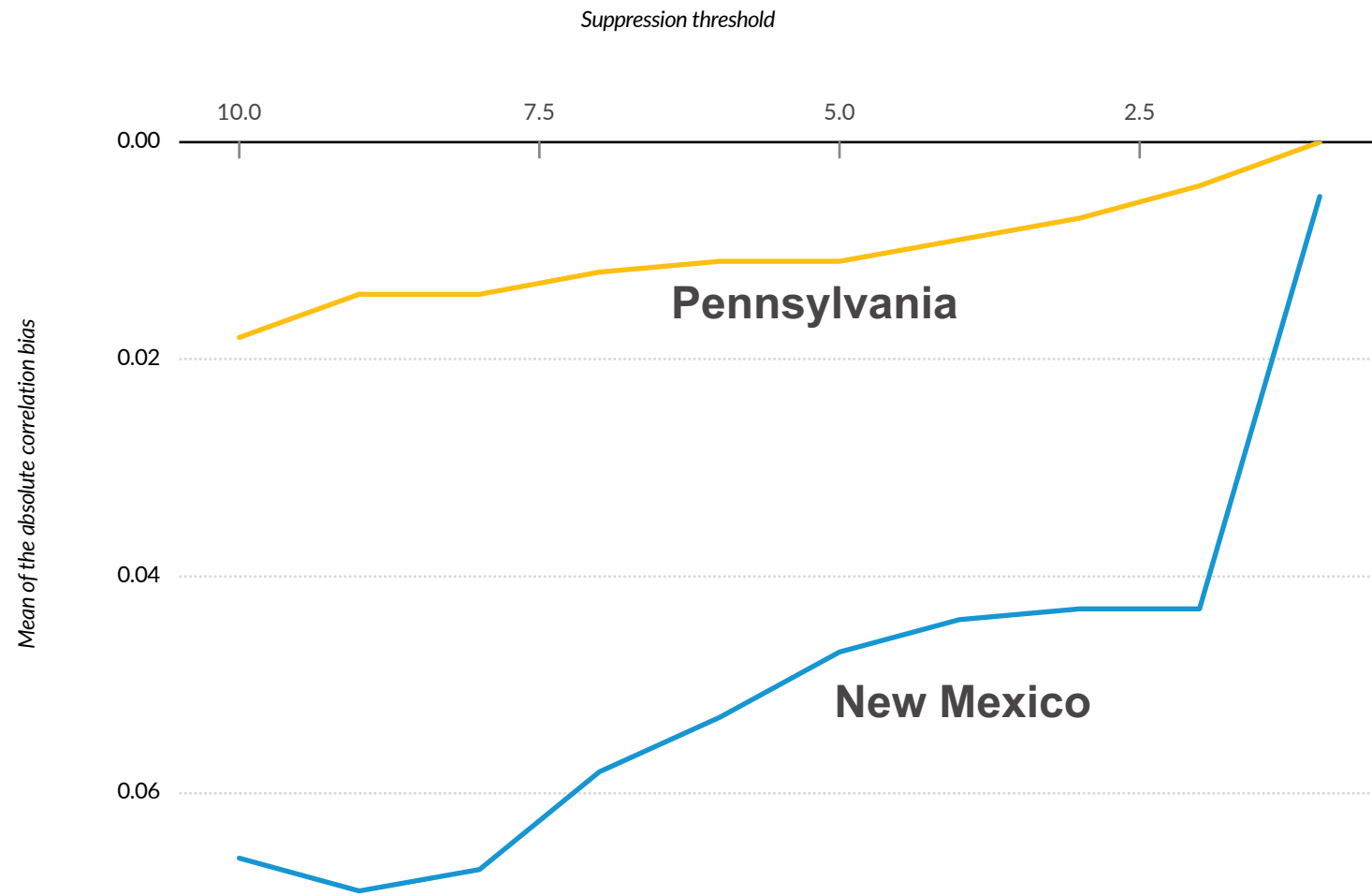
New Mexico



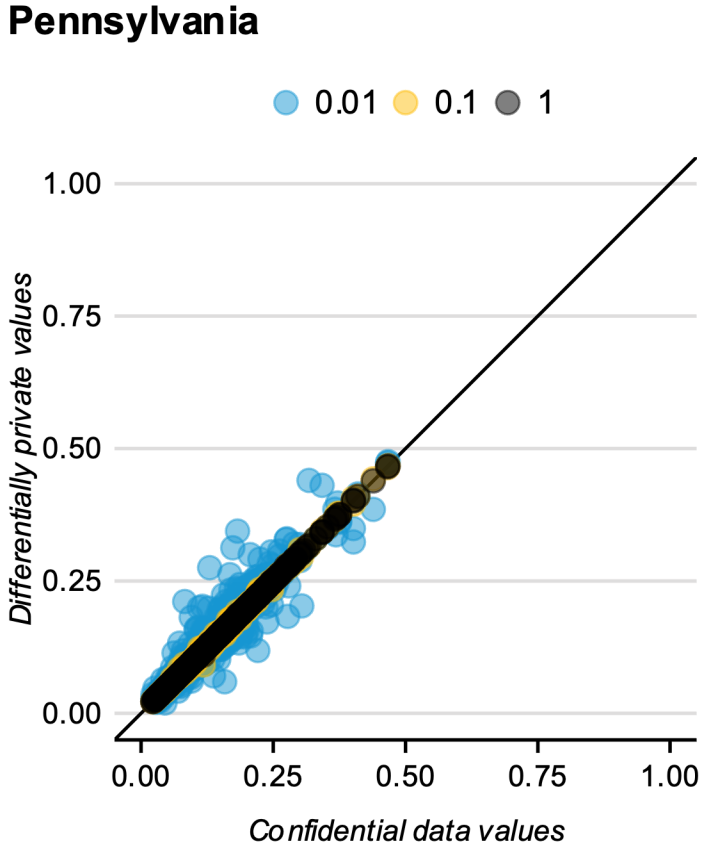
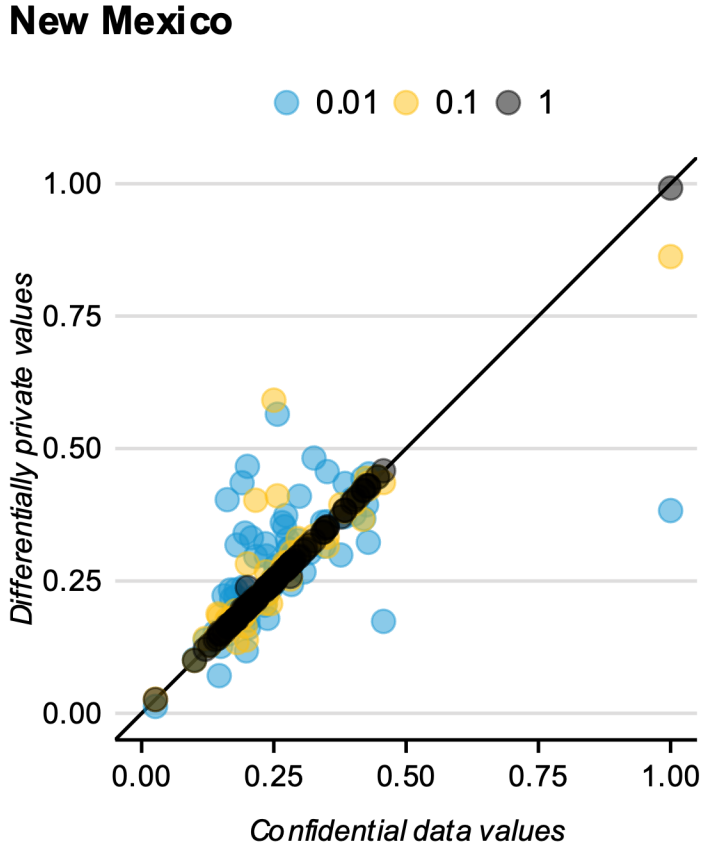
Pennsylvania



Differing Tradeoffs for Suppression



Differing Tradeoffs for Differential Privacy



2. Definitions and Preferences of Privacy and Utility

“To what extent do the privacy protections we provide through some system correspond to people’s real-world concerns and potential harms from breaches of privacy or confidentiality?”



– Adam Smith, Boston University

- Various groups may define privacy/utility differently to best fit their needs and expectations
- Groups may want different thresholds of privacy/utility
- How can groups place themselves in *equally satisfactory* positions?

3. Defining Groups and Representation

“If everyone in the United States were carbon copies of one another, we wouldn’t have a lot of discussions about utility or equity.”

– Rolando Rodriguez, U.S. Census Bureau



- How do we define groups in the data?
- Who represents their interests?
- How do we handle the sample size constraints?

3. Defining Groups and Representation

“

We want equal protection for all groups.
But we're more concerned about
sufficiently protecting [underrepresented
groups].

JÖRG DRECHSLER, DISTINGUISHED RESEARCHER AT THE INSTITUTE FOR
EMPLOYMENT RESEARCH IN GERMANY AND ASSOCIATE RESEARCH PROFESSOR
AT THE UNIVERSITY OF MARYLAND

DO NO HARM GUIDE: APPLYING EQUITY AWARENESS IN DATA PRIVACY METHODS



4. Resource Allocation

“Most [discretionary requests] are turned down. The people that work [at the statistical agency] always want to do these things, but again it’s the timing, the resources...most of them they feel like they can’t.”



– Daniell Toth, Bureau of Labor Statistics

- How do we balance agency mandated information releases vs. discretionary requests?
- How do we provide access to restricted data?
- How do we train people on how to properly use the data?

4. Resource Allocation



*“I’m talking about students for example... that can’t afford hiring a stat analyst [who] has the skills to do some of the data science kind of stuff. **How much do you help these people be successful in their job?** ... How far do you go with this to help someone out who can’t get their code straightened out or can’t write well?”*

– Tom Krenzke, Westat

5. Equity in Related Processes

“I think every county should have the right ability to analyze the statistics we are preparing and be useful.”

– Mauricio Ortiz, Bureau of Economic Analysis



- SDP is only a part of a much larger pipeline
- How can we minimize inequalities in data quality?
- What is an adequate starting place?

6. Misalignment with other Equity Efforts

“If [equity] isn’t considered in the design of the study for that small subgroup, then it’s not the main purpose of the study. Other parts will take priority. After all, what was the study designed for?”



– Saki Kinney, RTI

- We should consider impact of more granular data collection
- Incorporate privacy conversations early
- We should not treat equity and privacy as separate studies

7. Equity in the Research Field

“Certain research methods are a priori seen as legitimate or illegitimate depending on the community. But pluralist viewpoints work for other fields so we should be open to that and to more collaboration as well.”

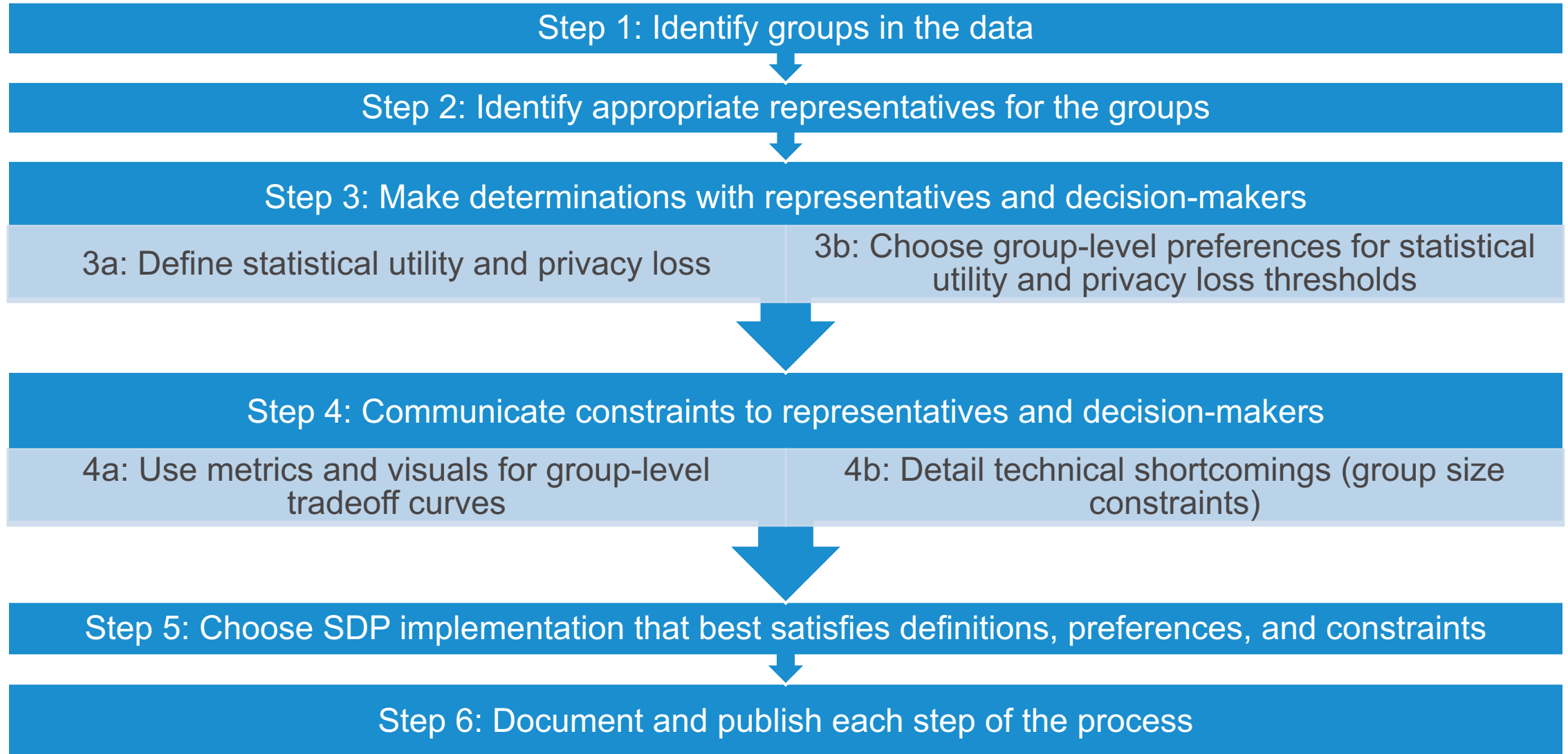


– Aleksandra Slavković, Penn State

- SDP researchers should collaborate more and incorporate other perspectives
- Increase openness to different approaches
- Acknowledge previous work

Where do we go from here?

An Aspirational Workflow for Equitable SDP



Concluding Recommendations

- Do not treat equity as a separate field of study
- Consider literature and perspectives from fields outside of your own
- Estimate separate privacy loss-utility curves for groups
- Work with groups represented in your data
- There is no methodological silver bullet

Concluding Recommendations

- Do not treat equity as a separate field of study
- Consider literature and perspectives from fields outside of your own
- Estimate separate privacy loss-utility curves for groups
- Work with groups represented in your data
- There is no methodological silver bullet

Thank you! Questions?

jsnoke@rand.org