The Case for Researching Applied Privacy Enhancing Technologies

Claire McKay Bowen[1], Joshua Snoke[2], Aaron R. Williams[1], and Andrés F. Barrientos[3]

[1] *Urban Institute, awilliams@urban.org & cbowen@urban.org*

[2] *RAND Corporation, jsnoke@rand.org*

[3] *Florida State University, abarrientos@fsu.edu*

Keywords— differential privacy, formal privacy, administrative data, policy analysis, linear regression, econometrics

Abstract: Research on privacy enhancing approaches for sharing data has grown significantly over the past two decades. This increased interest has led to extensive theoretical and methodological research, but the number of practical applications of privacy enhancing technologies has lagged far behind. This paper provides an overview of the Safe Data Technologies Project and the approach we have taken to conducting privacy research with the specific aim of putting theory into practice and incorporating user input. We provide an overview of the broader project goals, which aim to safely expand access to administrative tax data that is currently highly restricted. We highlight how understanding user interactions with the privacy enhancing methods has driven our research path and challenged the often unrealistic assumptions underlying much of the theoretical work. We review the primary findings from our research, discuss our plans for future directions, and make the case for researchers to pursue similar lines of applied inquiry.

1 The Value of Data Sharing for Policymaking

Accessing data, particularly administrative and survey data, is essential for improving evidence-based policymaking for government officials, policymakers, social science researchers, and data practitioners. For instance, Nagaraj and Tranchero (2023) demonstrated how direct access to confidential administrative data impacted the rate, direction, and policy relevance of economics research. One of their findings showed that researchers with confidential administrative data access are more likely to produce papers that receive more citations in public policy documents and produced 24% more publications in top journals per year. Echoing this sentiment, former Under Secretary for Economic Affairs in the Department of Commerce, Jed Kolko, wrote a blog post[1] that states one of the three types of useful research comes from papers focused on "...analyses that directly quantify or simulate policy decisions." Conducting relevant research for impactful policy work often requires accessing administrative data.

1.1 Privacy Challenges in Accessing Administrative Tax Data

Despite the immense value administrative tax data brings for evidence-based policymaking, the sensitivity of the data presents barriers to accessibility. Secure data access to administrative data is often restricted to select government agencies, a limited number of researchers working in collaboration with analysts in those agencies, and highly selective research programs run by these agencies. For example, if a researcher wanted access to U.S. taxpayer data, they must be a U.S. citizen before applying for a highly selective research program[2] through the Statistics of Income (SOI) Division at the IRS. If selected, the researcher would then undergo an extensive clearance process that could take several months before gaining access to the data at a secure data enclave.

For many researchers in the United States, they do not meet the eligibility requirements, may not be selected for the program, or may not pass the clearance process. This creates a small pool of researchers with access. For those outside this selective group, they may rely on public statistics or public data files; often referred to as public use files (PUF). SOI annually releases a PUF that is a privacy-protected database of sampled individual income tax returns. Yet, PUFs come with their own challenges. Privacy protections, such as aggregation, must be applied to balance the potential risks to respondents against the need for sharing the data.

In the case of administrative tax data, only a few trusted institutions have access to the SOI PUF (e.g., Urban-Brookings Tax Policy Center) and they must sign an MOU and pay $9,000[3]. Even for those institutions that have access to the SOI PUF, SOI has increasingly restricted and altered it over the years due to growing data privacy concerns (Bryant, Czajka, Ivsin, & Nunns,

[1] "The economic research policymakers actually need." Accessed August 12, 2024. https://www.slowboring.com/p/the-economic-research-policymakers

[2] "Statistics of Income Joint Statistical Research Program," Accessed August 12, 2024. https://www.irs.gov/statistics/soi-tax-stats-joint-statistical-research-program

[3] The fee helps covers the cost to create the file, which tends to be extremely labor intensive.

2014). Further, the statistical data privacy methods are time consuming for SOI staff and change the data's statistical properties in unknown ways, which reduce the data's usefulness for analysis.

Even researchers who are selected into the competitive program and pass the clearance process can only access the data at a secure data enclave or on a government issued laptop. If the former, these enclaves tend to be located at an academic institution in a highly populated area, so if the researcher is from a smaller institution with less resources, they may have to travel hundreds of miles to reach the closest secure data center.[4] The distance places a huge burden on the researcher to find time and funds to travel to the enclave. The result of all these challenges is that a substantial portion of economic and social science researchers will never gain access to the U.S. taxpayer data, and the differences reflect certain data access inequalities. Access is often granted to researchers who meet certain criteria: a U.S. citizen, part of a prestigious institution in certain geographic areas, and well-funded.

1.2 Developing Another Tier of Data Access

With the goal of safely expanding access to administrative tax data, our project team has been working with SOI to develop an automated validation server using formal privacy (Barrientos, Williams, Snoke, & Bowen, 2024; Taylor, MacDonald, Ueyama, & Bowen, 2021) and improve the SOI PUF using synthetic data generation (Bowen, Bryant, Burman, et al., 2022; Bowen, Bryant, et al., 2022; Bowen et al., 2020), as complements to the secure access program for the confidential data. We call this collaboration the Safe Data Technologies Project, which has involved over 30 researchers across five institutions since its formation. While the synthetic data effort comprises a crucial part of the project, we focus this article on the research efforts towards creating a validation server.

A *validation server* allows users to submit and run statistical queries on the confidential (i.e., validation) data after the users have developed their queries using the publicly released data. This represents one of the possible new tiers between secure data access to the confidential data and the public data releases, and it has been identified as a potential solution in multiple reports, such

[4] One of the authors for this paper is over 400 miles away from the closest federal statistical research data center.

as the Advisory Committee on Data for Evidence Building Year 2 Report[5] and Committee on National Statistics report series on “Toward a 21st Century National Data Infrastructure” (Reiter et al., 2024).

The validation server development poses several technical, practical, and policy challenges. In the following sections, we review a series of three papers which we wrote as we researched the practicality of creating an automated validation server using formal privacy, namely differential privacy. The papers moved in a progression from first testing the currently available methodology on *real-data* to collecting information from potential users about their needs and expectations when interacting with a privacy system before finally putting these results together to create a benchmarking framework for evaluating future methods. The key insights we produced in each study came from seeking to create a practical application, challenging the theoretical assumptions, and involving user perspectives.

2 Phase 1: Challenging Theory with Real-Data

With the goal of creating an automated validation server, we first researched how to produce a system to provide consistent and robust privacy protection with little or no human review. Differential privacy (DP), a concept proposed by Dwork, McSherry, Nissim, and Smith, 2006 which has gained substantial traction in the past two decades, provides an attractive option to automate the review process, removing the human element. At a high level, DP and related formal privacy definitions provide an a priori privacy guarantee which when applied consistently enables a specific type of automatic privacy accounting. Several definitions of different types of DP exist, and we refer to methods that satisfy one of these definitions as DP[6] methods. Satisfying DP is a provable feature of a method, not the data—a common misconception. See Williams and Bowen, 2023 for further mathematical details and review of DP and other formally private methods.

Evaluating the potential to use DP in an automated validation server, we conducted an extensive study on various state-of-the-art differentially private mechanisms (Barrientos, Williams, Snoke, & Bowen, 2024) to understand the feasibility of current DP methods for querying summary

[5] “Advisory Committee on Data for Evidence Building.” Accessed on April 30, 2024. https://www.bea.gov/evidence

[6] Note that we use DP as an acronym for both “differential privacy” and “differentially private”.

statistics and regression analyses. We chose queries that fit the use case of a validation server for tax policy research, drawing on input from our tax economist collaborators. In this paper, we tested methods for tabular statistics, mean statistics, quantile statistics, and statistics for full inference from a linear regression model with cross-sectional data. There are several other analyses we identified that we did not test, such as model selection, regression discontinuity, and kink designs. These methods are important for tax policy researchers, but we found the current DP methodology for these techniques is either in its early stages of development or does not support them at all.

2.1 A Feasibility Study of Differentially Private Methods

For summary statistics, we explored tabular statistics along with quantiles and means with their associated standard deviations and confidence intervals. For our testing, we *excluded* some methods because they:

- Require the user to set a prior bounds on the standard deviation and users likely will not know this in practice.
- Are highly sensitive to the assumption that the data are symmetric.
- Does not produce full inferences (i.e., methods that do not produce standard errors or confidence intervals).

For regression, we assessed DP regression methods with their associated standard deviations and confidence intervals. We selected methods if we could:

- Use the method for linear regression with normal errors, handle multiple predictors, and produce output that enabled full inference.
- Verify if the method achieves provable DP.
- Determine if the manuscript for a particular method provided all the required details for implementing the algorithm.
- If there were any issues in practically implementing the algorithm, provided a potential solution to address them.

2.2 Key Findings and Conclusions

Our work in Barrientos, Williams, Snoke, and Bowen, 2024[7] is the first comprehensive evaluation of these DP methods for practical applications within a validation server framework. We found that DP methods which provide summary statistics like means and percentiles perform well, whereas obtaining full inference on DP regression coefficients performs poorly. The latter result came as a surprise given many privacy experts may have believed that DP linear regression was a solved problem due to the numerous papers and these methods' reasonable large sample properties. However, in Barrientos, Williams, Snoke, and Bowen, 2024 we found only one method that met all the inclusion criteria without additional adaptations (Ferrando, Wang, & Sheldon, 2021). We included one other method, because even though it was not originally designed for linear regression, we made an adaptation to it and made it eligible (Brawner & Honaker, 2018). We eventually increased the number of testable methods from two to six by re-purposing elements of the algorithm from Ferrando, Wang, and Sheldon, 2021 to enable full inference with other mechanisms. Despite making these changes, the tested methods performed poorly due to either inflating the confidence intervals so severely as to limit any conclusions that could be drawn from the data, or the output did not appropriately account for the uncertainty and led to erroneous inferences.

In addition to the methodological issues, we encountered several challenges with coding the various algorithms. We do not expect academics to provide production-ready code, but code we obtained from them often fell short of standards for reproducibility, such as those set by the American Economic Association (AEA)[8]. We frequently found the research code to be messy, hard to read, and difficult to alter for our use cases. Most manuscripts describing the methods do not provide enough information to properly implement the method, which resulted in exclusion from the feasibility study. These barriers in implementing other research methods stress the importance of providing open-source code to facilitate wider use and acceptance of DP algorithms in practice.

[7] All results from the paper can be replicated using the code on GitHub at https://github.com/UrbanInstitute/formal-privacy-comp-appendix.

[8] American Economic Association's Data and Code Availability Policy. Accessed: 2024-07-25. https://www.aeaweb.org/journals/data/data-code-policy

3 Phase 2: Connecting Empirical Results with Users' Input

In Barrientos, Williams, Snoke, and Bowen, 2024, we identified which summary statistics and regression DP methods perform the best against each other, but we did not identify thresholds to determine if the formally private outputs are accurate enough for public policy decisions. In other words, the data privacy community generally focuses on the trade-off between accuracy and privacy, but we lack absolute standards for how accurate results must be to inform policy decisions. Put another way, we need to know the required level of accuracy at which users would find the validation server useful or viable. The lack of a standard or threshold presents a significant hurdle to implementing a DP validation server in practice.

3.1 An Assessment of Economists' Understanding, Perception, and Tolerance for Formal Privacy

In an attempt to *start* answering these questions in a novel way, we conducted a convenience sample survey of members of AEA to evaluate and identify the expectations and needs of potential users for a validation server (Williams, Snoke, Bowen, & Barrientos, 2024). Our survey questions aimed to identify the baseline knowledge of economists about DP, their attitudes toward DP frameworks, the types of statistical methods that they think are most useful, their tolerance for privacy induced errors, and how they would spend their privacy loss budget. Through this study, we gained some of the first in-depth insights regarding economists' perspectives and opinions' concerning formal privacy, and we introduced a framework for surveying potential validation server users.

The questionnaire began with asking the participant/economist about their demographic and professional characteristics. These questions allowed us to determine if the respondents to our questionnaire resemble our population of interest and to compare responses across demographic groups. The next section of the questionnaire asked about the types of methods research economists use with cross-sectional data. The questionnaire then evaluated research economists' knowledge and perceptions of formal privacy and DP. For the final section of the questionnaire, we included vignettes to explore research economists' tolerance for errors from DP and their preferences for using DP. Vignettes can approximate real-world behavior by presenting respondents with competing choices (Hainmueller, Hangartner, & Yamamoto, 2015).

We collected our data from members of AEA, a professional organization of about 23,000 professionals or graduate-level students dedicated to economics research and teaching[9]. EA sent the email to 8,850 economists (who had opted-in to surveys). We did not offer any incentives for completing the questionnaire. We received a large response of over 1,000 individuals.[10]

3.2 Key Findings and Conclusions

Based on our survey results, we learned that economists have a limited understanding of DP and formal privacy. For example, survey design research suggests using prominent events to elicit more accurate responses, such as "before the COVID-19 pandemic" or "after 9/11" (Tourangeau, Rips, & Rasinski, 2000). In this vein, our survey ask respondents how many people in their professional circles have discussed the U.S. Census Bureau's adoption of DP/formal privacy for the 2020 Decennial Census. This change in the Census Bureau's disclosure avoidance system affected a major source of data for empirical research and spawned widespread debate (Ruggles & Van Riper, 2022) and popular news coverage (Bahrampur & Lang, 2021; Wang, 2021; Wines, 2022). Despite being asked about the highly debated adoption of DP by the U.S. Census Bureau, a significant majority of respondents (68.3%) reported that they did not know of anyone in their professional circles who discussed it. Our survey also identified economists' mixed skepticism about DP and formal privacy. This suggests that our project team has work to do to thoroughly motivate a formally private validation server along with providing substantial training on its usage and how users should report the noisy results.

For desired methods and analyses, the survey results showed that economists want the ability to merge multiple data sets (i.e., data blending). Although data blending is important for economists and other social scientists, formal privacy literature rarely discusses it. Another finding is that our respondents are interested in a wide range of econometric methods, such as panel data methods and difference-in-difference, allowing for more sophisticated research designs in addition to multiple linear regression on a single cross-section of data. Despite the demand for this type of analyses, such methods are almost non-existent in the formal privacy literature.

[9] For more information about the American Economic Association, see their website at https://www.aeaweb.org/about-aea

[10] GitHub repo website, https://github.com/UrbanInstitute/formal-privacy-aea-questionnaire

In the vignettes section, our respondents generally had low tolerances for errors when posed with the trade-offs between errors and adversely responding as journal referees. In addition to improving communication about formal privacy and usage of a validation server, more work is needed for peer-reviewed journals to accept outputs from a formally private validation server system. We suspect that without strong professional incentives related to academic publishing, most researchers may not invest the effort needed to learn how to use a new data access system.

After conducting this study, we encourage other privacy researchers to benchmark their methods against users' error tolerances and expectations instead of what is typically done; benchmarking against other formally private methods and seeing which uses the least amount of privacy loss budget for a prediction output. Although researchers understand the importance of accessing administrative data, our survey indicated they would rather sacrifice such access than have results from a formally private system if the errors are too high.

4 Phase 3: A Practical Approach to Benchmarking Formally Private Methods

Our finding from Barrientos, Williams, Snoke, and Bowen, 2024 and Williams, Snoke, Bowen, and Barrientos, 2024 suggested that current DP linear regression methods are unlikely to support full regression-based inference on an administrative tax data validation server or may only be utilized at high costs to the privacy budget. We see two main reasons exist for this gap between theoretical expectation and empirical reality. First, we have a finite, often small, sample size when working with real-data. This gap has already been recognized as an issue in the DP literature (Slavković & Seeman, 2023) with some methods designed specifically for statistical inference under finite samples in certain cases (Awan & Slavković, 2018; Vu & Slavkovic, 2009).

Second, and perhaps more substantially, the simulation studies conducted in papers proposing new DP mechanisms only consider situations where the assumptions of ordinary least squares (OLS) are satisfied and the coefficients have a strong signal. This is not the case in many applications of linear regression for economic, statistical, and social science research. Some examples are the residuals may be skewed or heteroscedastic, there may be multicollinearity between predictor variables, or categorical variables may be imbalanced. We have not found any prior work which considers the interaction of adding noise to satisfy DP for OLS models where one or more of these violations exist.

These findings led us to propose a framework for explicitly testing DP mechanisms under different scenarios, so that we can better understand how existing mechanisms will work (or not work) when applied to real-data (Williams, Barrientos, Snoke, & Bowen, 2024). To create a framework for empirically benchmarking DP linear regression methods under different real-data scenarios, we built directly from past studies (Barrientos, Williams, Snoke, & Bowen, 2024; Williams, Snoke, Bowen, & Barrientos, 2024) to develop a simulation framework that systematically explores the accuracy and precision of performing full inference using the output from DP regression methods for multiple linear regression.

4.1 Simulation Framework Design

Our empirical study considers the accuracy and precision of DP estimates for performing full inference with a regression model under a variety of different settings. We summarize the accuracy and precision using utility metrics such as the relative absolute error, the effective sample size (i.e., relative to the no-noise model), and the coverage rates.

We tested the best performing DP regression methods identified in Barrientos, Williams, Snoke, and Bowen (2024), which are the Laplace mechanism (Ferrando, Wang, & Sheldon, 2021) and the Analytic Gaussian mechanism (Balle & Wang, 2018). We then used the results from Williams, Snoke, Bowen, and Barrientos (2024) to benchmark users' expectations and error tolerances on the DP outputs. The purpose of benchmarking against users' expectations and error tolerances is to provide a practical bar against which any viable method must pass.

We grouped the various factors that can influence a regression analysis into three categories. The first group pertained to the data-generating distribution and does not imply a violation of the model assumptions. For instance, this category included changes in the signal-to-noise ratio (SNR), probabilities of observing specific categories in categorical variables, correlations among continuous variables, and choices of reference levels in categorical variables. The second group included violations of model assumptions, such as non-normally distributed errors and non-constant error variance. The last group related to the input parameters necessary for implementing DP, such as specifying variable ranges to bound global sensitivity, the privacy budget, and the noise injection mechanism. The first two groups are issues that can affect OLS estimates without DP, while the last group relates specifically to the application of DP.

We empirically study the influence of these factors through a simulation study. While we acknowledge the limitations of this approach, empirical studies can shed light on key issues where the theoretical assumptions do not hold. They aid in the prioritization of investigation for theoretical gaps. Empirical studies can also help practitioners understand which aspects of the data and implementation require caution when employing DP in regression analysis. Furthermore, this empirical study can serve as an example and provide a framework for assessing the influence of such factors when privacy researchers introduce new DP regression approaches.

In the first stage of our simulation framework, we evaluated the performance of the DP mechanisms under *favorable* conditions regarding the data generation mechanism. Specifically, we generated data from an underlying generating distribution that does not violate any of the modeling assumptions. In the second stage, we assessed the performance of the approaches under multiple alternative scenarios characterized by violation of assumptions, multicollinearity, and categorical covariates with low-frequency levels.

Finally, we leveraged findings from both our simulation study and the survey conducted in Williams, Snoke, Bowen, and Barrientos, 2024. Using the tolerance levels identified in Williams, Snoke, Bowen, and Barrientos, 2024, we determined the scenarios under which candidate DP methods would meet users' expectations. Our results provide an example of how users' input could be combined with empirical studies to test the practicality of DP methods. Future work could and should utilize other information gatherered from relevant users.

4.2 Key Findings and Extensions

In Williams, Barrientos, Snoke, and Bowen, 2024, we provided the infrastructure of an assessment framework for future DP regression methods. Privacy researchers can use this framework to assess new theoretical ideas under a variety of real-data scenarios and privacy budgets, with a focus on statistical inferences rather than predictions. Potential users of a validation server who can also utilize our assessment framework to formulate an analysis plan, similar to a power analysis, on a synthetic data set before accessing a validation server that produces formally private outputs. Our code is available online[11] for anyone interested in using this evaluation framework.

Another key finding is that researchers interested in evaluating DP methods for regression via simulation studies must carefully consider how to set bounds for continuous unbounded variables. Williams, Barrientos, Snoke, and Bowen, 2024 showed how different bounds significantly impact method performance. Therefore, we recommend that simulation studies for regression analysis should involve unbounded continuous variables and systematically test the robustness of results under various strategies for setting up bounds.

In general, the DP regression models we tested did not perform well for inference unless the SNR is high and all the assumptions of OLS are satisfied, and the privacy budget is moderate to small. It is striking that the quality of the results varied significantly with relatively modest changes to the scenarios. For example, modest changes to the approach for setting bounds on the data resulted in wildly different effective sample sizes. Outside of a simulation environment, it will be difficult for analysts and users of DP tools to anticipate the exact impact of DP noise on linear regression results, but the simulations help users get a much better idea than what theory might suggest.

In our paper, we only explored a small subset of possible alternative scenarios. Future research will extend this framework to consider other violations of normal linear model assumptions, such as non-linearity, independence of errors, and zero conditional mean assumption. Additionally, we will explore aspects like omitted variable bias, varying sample sizes and the number of predictors, incorporating different regularization strategies, and considering different privacy budgets to establish DP bounds for continuous unbounded variables. Moreover, we will introduce new utility metrics, such as the coverage of predictive intervals, to provide a more comprehensive evaluation framework.

5 Concluding Thoughts for Future Research

In Snoke, Bowen, Williams, and Barrientos, 2023, we highlighted lessons learned and roadblocks encountered throughout the Safe Data Technologies project. In our efforts to conduct applied data privacy research, we repeatedly find that the problems encountered come from fundamental con-

[11] GitHub repo website is forthcoming.

flicting assumptions. We refer to them as *incompatibilities* between current practices in statistical data analysis and statistical data privacy, which are: 1) estimates for traditional statistical inference; 2) control or nuisance variables; 3) assumptions on the range of the data or other assumptions; and 4) performing exploratory data analysis. We discussed how overcoming these incompatibilities will require many compromises and changes within the data privacy community in how we approach statistical data analysis and statistical data privacy together. From our previous studies and reflecting on these challenges, we outline a few areas of possible future work.

One vital area for improvement is developing DP algorithms for data that are not well-behaved, i.e., Gaussian. Another area of DP method development is enabling full inference from statistical models and working with finite sample sizes similar to what is utilized in social science or economic research. Additionally, Williams, Snoke, Bowen, and Barrientos (2024) showed there is a demand for many other methods that can address more sophisticated research designs, such as regression discontinuity, and kink designs.

Any new DP method should be tested using a framework that simulates real-world data. Williams, Barrientos, Snoke, and Bowen (2024) provided a framework for ensuring DP linear regression methods are useful to social scientists who focus on inference. The idea is the benchmarking tool would resemble power analysis but for statistical data privacy methods, and it would be particularly useful for users of a validation server who do not have access to the confidential data.

DP methods also must be benchmarked against users' expectations and error tolerances to provide a practical bar against which any viable method (DP or not) must pass. In our work, we contributed to the literature in giving a measuring bar for our benchmarking framework and others that is explicitly tied to users' expectations (Williams, Barrientos, Snoke, & Bowen, 2024). In only one survey we conducted, we learned a substantial amount about economists' perspectives and their likelihood of using a validation server under different constraints (Williams, Snoke, Bowen, & Barrientos, 2024). We hope others will use the lessons learned from our study to conduct future surveys to expand our knowledge base about users. Repeated surveys of different user groups would improve on our benchmarking estimates and help inform what what methods are most needed for the entire data user community.

In summary, our research on applied statistical data privacy methods has led to several novel discoveries and insights that would not have emerged from theoretical research alone. In particular, the use of real-world data and real statistical analyses that social scientists utilize enabled us to identify the gaps between the theoretical and empirical performance of existing DP methods for linear regression. Also by directly surveying non-privacy researchers, we gathered useful insight into the types of methods which need to be developed to meet users' needs. We showed how to create a benchmark for setting privacy budgets that is based on users' accuracy expectations. These findings led us to create a simulation framework for testing and evaluating DP methods that we hope will help move the field closer to developing practical privacy applications.

Acknowledgments

This research was funded by the Alfred P. Sloan Foundation [G-2022-17149] and National Science Foundation National Center for Science and Engineering Statistics [49100422C0008].

We would like to thank our collaborators at SOI, especially Barry Johnson, Victoria Bryant, Chris Rexrode, Conrado Arroyo, Derek Gutierrez, and Giang Trinh for their amazing support.

We also thank our stellar Safe Data Technologies Project team, consisting of Nikhita Airi, Leonard Burman, John Czajka, Surachai Khitatrakun, Graham MacDonald, Rob McClelland, Sybil Mendonca, Josh Miller, Gabriel Morrison, Liudas Panavas, Jeremy Seeman, Jean Clayton Seraphin, Deena Tamaroff, Silke Taylor, Erika Tyagi, and Doug Wissoker.

Finally, we thank our advisory board for their advice and support. The members are John Abowd, Jim Cilke, Connie Citro, Jason DeBacker, Rick Evans, Dan Feenberg, Max Ghenis, Nick Hart, Matt Jensen, Ithai Lurie, Ashwin Machanavajjhala, Shelly Martinez, Robert Moffitt, Amy O'Hara, Mauricio Ortiz, Nancy Potok, Jerry Reiter, Rolando Rodriguez, Emmanuel Saez, Wade Shen, Aleksandra Slavković, Salil Vadhan, and Lars Vilhuber.

Conflict of Interest

The authors report there are no competing interests to declare.

References

Awan, J., & Slavković, A. (2018). Differentially private uniformly most powerful tests for binomial data. *Advances in Neural Information Processing Systems*, *31*.

Bahrampur, T., & Lang, M. J. (2021). New system to protect census data may compromise accuracy, some experts say. *The Washington Post*.

Balle, B., & Wang, Y.-X. (2018). Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising, 394–403.

Barrientos, A. F., Williams, A. R., Snoke, J., & Bowen, C. M. (2024). A feasibility study of differentially private summary statistics and regression analyses with evaluations on administrative and survey data. *Journal of the American Statistical Association*, *119*(545), 52–65.

Bowen, C. M., Bryant, V., Burman, L., Czajka, J., Khitatrakun, S., MacDonald, G., McClelland, R., Mucciolo, L., Pickens, M., Ueyama, K., et al. (2022). Synthetic individual income tax data: Methodology, utility, and privacy implications. *International Conference on Privacy in Statistical Databases*, 191–204.

Bowen, C. M., Bryant, V., Burman, L., Khitatrakun, S., McClelland, R., Stallworth, P., Ueyama, K., & Williams, A. R. (2020). A synthetic supplemental public use file of low-income information return data: Methodology, utility, and privacy implications. *International Conference on Privacy in Statistical Databases*, 257–270.

Bowen, C. M., Bryant, V. L., Burman, L., Khitatrakun, S., McClelland, R., Mucciolo, L., Pickens, M., & Williams, A. R. (2022). Synthetic individual income tax data: Promises and challenges. *National Tax Journal*, *75*(4), 767–790.

Brawner, T., & Honaker, J. (2018). Bootstrap inference and differential privacy: Standard errors for free [unpublished manuscript].

Bryant, V. L., Czajka, J. L., Ivsin, G., & Nunns, J. (2014). Design changes to the soi public use file (puf). *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, *107*, 1–19.

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography* (pp. 265–84). Springer.

Ferrando, C., Wang, S., & Sheldon, D. (2021). General-purpose differentially-private confidence intervals. *arXiv preprint arXiv:2006.07749*.

Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, *112*(8), 2395–400.

Nagaraj, A., & Tranchero, M. (2023). *How does data access shape science? evidence from the impact of us census's research data centers on economics research* (tech. rep.). National Bureau of Economic Research.

Reiter, J., Bowen, C., Cohen, A., Farrell, D., Goerge M., R., Hart, N., Hosagrahar V., J., Kifer, D., Levy, K., Viljoen, S., & Watson, M. (2024). Toward a 21st century national data infrastructure: Managing privacy and confidentiality risks with blended data. *National Academies of Sciences, Engineering, and Medicine.*

Ruggles, S., & Van Riper, D. (2022). The role of chance in the census bureau database reconstruction experiment. *Population Research and Policy Review*, *41*, 781–788.

Slavković, A., & Seeman, J. (2023). Statistical data privacy: A song of privacy and utility. *Annual Review of Statistics and Its Application*, *10*, 189–218.

Snoke, J., Bowen, C. M., Williams, A. R., & Barrientos, A. F. (2023). Incompatibilities between current practices in statistical data analysis and differential privacy. *arXiv preprint arXiv:2309.16703*.

Taylor, S., MacDonald, G., Ueyama, K., & Bowen, C. (2021). A privacy-preserving validation server prototype.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* Cambridge University Press.

Vu, D., & Slavkovic, A. (2009). Differential privacy for clinical trial data: Preliminary evaluations. *2009 IEEE International Conference on Data Mining Workshops*, 138–143.

Wang, H. L. (2021). For the u.s. census, keeping your data anonymous and useful is a tricky balance. *NPR*.

Williams, A. R., Barrientos, A. F., Snoke, J., & Bowen, C. (2024). Benchmarking DP Linear Regression Methods for Statistical Inference. *National Bureau of Economic Research*.

Williams, A. R., & Bowen, C. M. (2023). The promise and limitations of formal privacy. *Wiley Interdisciplinary Reviews: Computational Statistics*, *15*(6), e1615.

Williams, A. R., Snoke, J., Bowen, C., & Barrientos, A. F. (2024). Disclosing economists' privacy perspectives: A survey of american economic association members on differential privacy and data fitness for use standards. *Harvard Data Science Review*.

Wines, M. (2022). The 2020 census suggests that people live underwater. there's a reason. *The New York Times*.