Reproducibility of Social Science Research Under Privacy Noise

Ryan Steed*[1], Eduardo Abraham Schnadower Mustri[1], and Alessandro Acquisti[2]

[1]*Carnegie Mellon University*
[2]*Massachusetts Institute of Technology*

December 10, 2025

Abstract

Privacy-preserving analytics such as differential privacy are designed to allow the analysis of sensitive datasets while protecting individuals' privacy. Their deployment, however, has been controversial. Critics maintain that statistical noise injected to preserve privacy can degrade quality and feasibility of social science research [1, 2]. We select a benchmark of empirical results from 93 published social science studies and evaluate whether their findings replicate on privacy noise-infused datasets. Under privacy budgets typical in industry, around 91% of simulated findings still support the original claims at confidence level $\alpha = 0.1$. Claims based on weaker original effect sizes are more likely to be nullified or sometimes reversed. We compare distortions caused by privacy noise to those due to measurement errors and other kinds of non-sampling errors common in social statistics, and we find that the marginal impacts of privacy protection are smaller. Moreover, discrepancies due to privacy noise are often much smaller than discrepancies observed in traditional replication and robustness studies [3–5]. We point towards robust estimation techniques that can better account for noise—from data privacy protections or from data error—and enable social science research that is both rigorous and privacy-preserving.

*ryansteed@cmu.edu

Industry and government organizations from Meta [6] to the U.S. Census Bureau [7] have turned to new methods to strengthen privacy protections in data releases following concerns about attacks aided by commercial data and increased computing power [8, 9].[1]

Applications of these modern privacy protections, however, have met with controversy [18]. One of the most popular approaches, differential privacy (DP) [19, 20], is usually accomplished by injecting statistical noise when sharing statistics about sensitive personal data.[2] As new policies—including presidential executive orders—incentivize the adoption of DP and similar methods [25, 26], some researchers, policymakers, and civil rights advocates have worried that noise injected for privacy could result in a "precipitous decline in the quantity and quality of evidence-based policy research" [2], "greatly reduced social knowledge" [1], and distorted government decision-making [27, 28]. On the other hand, other scholars, public officials, and privacy advocates argue that *failing* to deploy modern protections will forfeit opportunities for research on previously inaccessible private data and potentially deteriorate public trust and future data quality [8, 29, 30]. To date, however, there is little systematic evidence to inform this debate: quantitative estimates of the impact of modern privacy protections on the outcomes of social science research are scant.

We identified, collected, and replicated a benchmark of 303 findings from 93 empirical studies published in economics and social science journals, and then compared their published results to a counterfactual version simulated with differentially private mechanisms. We evaluated *epistemic parity*, the principle that published findings should be replicable with a noise-infused dataset [31]. Would social science studies reach the same conclusions if they relied on statistics infused with noise for privacy?

We find that at the scale of privacy budgets typical in industry, around 91% of simulated findings still support the original claims at confidence level $\alpha = 0.1$; around 9% do not. At the scale of stronger privacy budgets, around 77% of findings still support the original claims. Notably, the rates of epistemic parity resulting from typical industry privacy budgets are generally higher than rates of parity in traditional large-scale replication studies in economics and psychology [3–5].

We analyze the antecedents of disparities due to additive noise, quantifying the impacts of privacy budget, choice of mechanism, and other characteristics of the statistics and analysis. We show that claims based on weaker initial effect sizes are more likely to be contradicted by added noise, as are more quantitatively specific claims.

[1]There are many empirical examples of reconstruction and re-identification attacks against public datasets, including attacks on Massachusetts Group Insurance Commission statistics [10], the Personal Genome project [11], the Netflix Prize dataset [12], the Aircloak Challenge [13], and, most recently, the 2010 and 2020 Decennial Censuses [14–17].

[2]Differential privacy has been adopted by Meta [6], the Wikimedia Foundation [21], the U.S. Census Bureau [7], and many other organizations [22–24].

We also compare the distortions caused by privacy noise to those due to measurement errors and other kinds of non-sampling errors common in social statistics [32, 33]. We find that even modest amounts of existing data error can alter findings at even higher rates. Differentially private mechanisms do exacerbate these existing disparities—but the marginal impacts of privacy protection are small.

We build on several studies examining the trade-offs between privacy and usability in public statistics, particularly in the context of the U.S. census [33–38]. Existing studies on research impacts examine synthetic microdata [31] and DP linear regression [39], methods which may be critical in future microdata products but are not yet widespread. Our study examines, at scale, the simpler and more common practice of infusing noise into aggregate statistics before public release, used in the 2020 Decennial Census [7] and the Social Science Facebook URLs dataset collaboration [6]. Our method of analysis is also inspired by replication efforts in psychology and economics, particularly the Open Science Collaboration [3, 4] and the Many Labs replication project [40].

Recent federal policy calls for the advancement of privacy-preserving data sharing across public and private sectors [25, 26]. The impacts of these measures would be widespread. The Integrated Public Use Microdata Series (IPUMS) alone, for example, lists over 27,000 academic publications based on its catalogue of U.S. census and survey data [41]; the Surveillance, Epidemiology, and End Results (SEER) program claims more than 17,000 publications use its survey data [42]; and thousands more studies rely on other public statistics published by government agencies and other organizations. Scholars have frequently called for more research into the potential impacts of differential privacy and similar techniques on research [1, 29, 43]. Our study provides an estimation of the key trade-offs involved and could inform policymakers and data curators designing the next generation of privacy-preserving technologies.

1 Building a large benchmark of reproduced studies

We used the *Inter-University Consortium for Political and Social Research (ICPSR)*'s replication package repository [44] and *Find Economic Articles with Data*, a tool for searching English economics articles with accessible data/code supplements [45], to find social science studies with working replication packages. These databases include articles from journals with standardized data accessibility policies, including the *American Economic Review*, the *American Economic Journal*s, and several other journals (SM Figure C.1). We searched for studies from the last two decades involving aggregate statistics about individuals or households (see §A.1) and successfully reproduced 93.

For each article, we read the abstract and introduction and identified the key empirical claims made by the author. We then identified the statistical estimate or estimates (hereafter *results*) supporting

Table 1: Descriptive statistics.

	Observations	Mean	SD	Min.	Q1	Median	Q3	Max.
Studies	**93**							
Years since publication	93	7.52	4.67	0	4.00	7.00	11.0	19.0
Num. results replicated	93	3.26	1.86	1.00	2.00	3.00	4.00	10.0
Citations (Semantic Scholar)	92	182.	530.	0	10.0	46.0	115.	3610.
Results	**303**							
Model degrees of freedom	292	60.3	131.	1.00	9.00	20.0	40.5	823.
Num. vars about people	303	8.20	12.6	1.00	3.00	5.00	10.5	110.
Proportion of personal vars. privatized	303	0.842	0.238	0.167	0.775	1.00	1.00	1.00
Answers primary or co-primary RQ	54							
Supports claim mentioned in abstract	269							
Personal data variables	**1026**							
Successfully privacy protected	859							
Role in regression								
Control var.	707							
Dependent var.	152							
Primary independent var.	67							
Instrumental var.	19							
Instrumented var.	27							
Subsetting var.	8							
Weighting var.	38							
Type of statistic								
Ratio	32							
Booleans/categorical	56							
Count, log count	146							
Mean, log mean	517							
Median	22							
More complex query	242							

Variables may be used in multiple results within the same study.

those claims—in our sample, all regression coefficients—and modified the authors' code to extract those key numerical results and related statistics (standard error, t-statistic, degrees of freedom, and so forth). For most studies, we reproduced 2–4 results (Table 1). Most results we reproduced (269 of 303) substantiate a key empirical claim mentioned in the article's abstract. Before noise injection, all 303 results we reproduced exactly matched the original results as published.

To simulate differentially private data release, we injected noise into all statistics about personal data used to produce the key results in our sample. We compared two common noise injection mechanisms (see §A.3): 1) pure differential privacy with the Laplace mechanism [20], a simple randomized mechanism for releasing statistics; and 2) zero-concentrated differential privacy (zCDP) [46, 47] with the Gaussian mechanism, designed to improve utility with a more relaxed definition of DP. In differential privacy, the parameters ϵ (pure DP) and ρ (zCDP) measure how much the inclusion of information about an individual in the dataset increases the risk that their personal information is revealed by published statistics (Definition 1).

2 Epistemic parity

Following Rosenblatt *et al.* [31], we test for *epistemic parity*: the principle that empirical claims should be replicable with a differentially private dataset.[3] A *claim* is an epistemic assertion based on one or more findings [31, 48] and a *finding* is a comparison between a statistical estimate (a result) and one or more other statistical estimates or scalars. For each finding, we define a region of possible coefficient values such that the original finding holds. For example, Guidetti *et al.* [49] claim that exposure to the pollutant PM10 increases pediatric hospitalization rates for respiratory disease based on the finding that the coefficient on PM10 exposure is greater than zero in a regression of hospitalization rates.

Strict epistemic parity is met when all the findings supporting a given claim have confidence intervals within this region at a given confidence level α. For example, most authors' claims require only that a statistically significant coefficient have a particular sign; others rely on particular relationships between the magnitude of different coefficients; a few rely on coefficient values to fall within a particular range.[4] (See SM Figure C.5 for examples.) We also test other common metrics of epistemic parity that do not depend on our interpretation of the authors' claims (Section A.2): *sign parity* and *sign reversal* (whether the direction of effect matches or does not match the original, respectively, at confidence level α) and *confidence interval (CI) coverage* (whether the confidence interval contains the original estimate at confidence level α).

3 Impacts of noise for differential privacy

Figure 1 presents the impacts of noise injected for differential privacy on several standardized metrics of epistemic parity developed in prior replication studies [4, 31, 39]. Additional metrics at additional confidence levels (effective sample size and absolute difference) are depicted in SM Figures C.2–C.3.

Though differential privacy advocates often prefer $\epsilon < 1$ [50], real-world public datasets released with differential privacy tend to use values of ϵ greater than 1 and sometimes even greater than 10. For example, Google's Community Mobility Reports use $\epsilon = 2.64$ per day, Facebook's URLs dataset was released with ($\epsilon = 1.453$, $\delta = 10^{-5}$)-DP, and the 2020 U.S. Decennial Census Demographic and Housing Characteristics tables used $\rho = 4.96$, equivalent to ($\epsilon = 26.34$, $\delta = 10^{-10}$)-DP [22–24, 51].

[3] We use the term epistemic *dis*parity to describe cases where epistemic parity does not hold.

[4] We except quantitative claims not originally based on statistical significance (e.g., a claim that an estimate fall between 0.1 and 0.2 despite its original confidence interval exceeding those bounds, or a claim that the one estimate exceeds another despite overlapping confidence intervals).

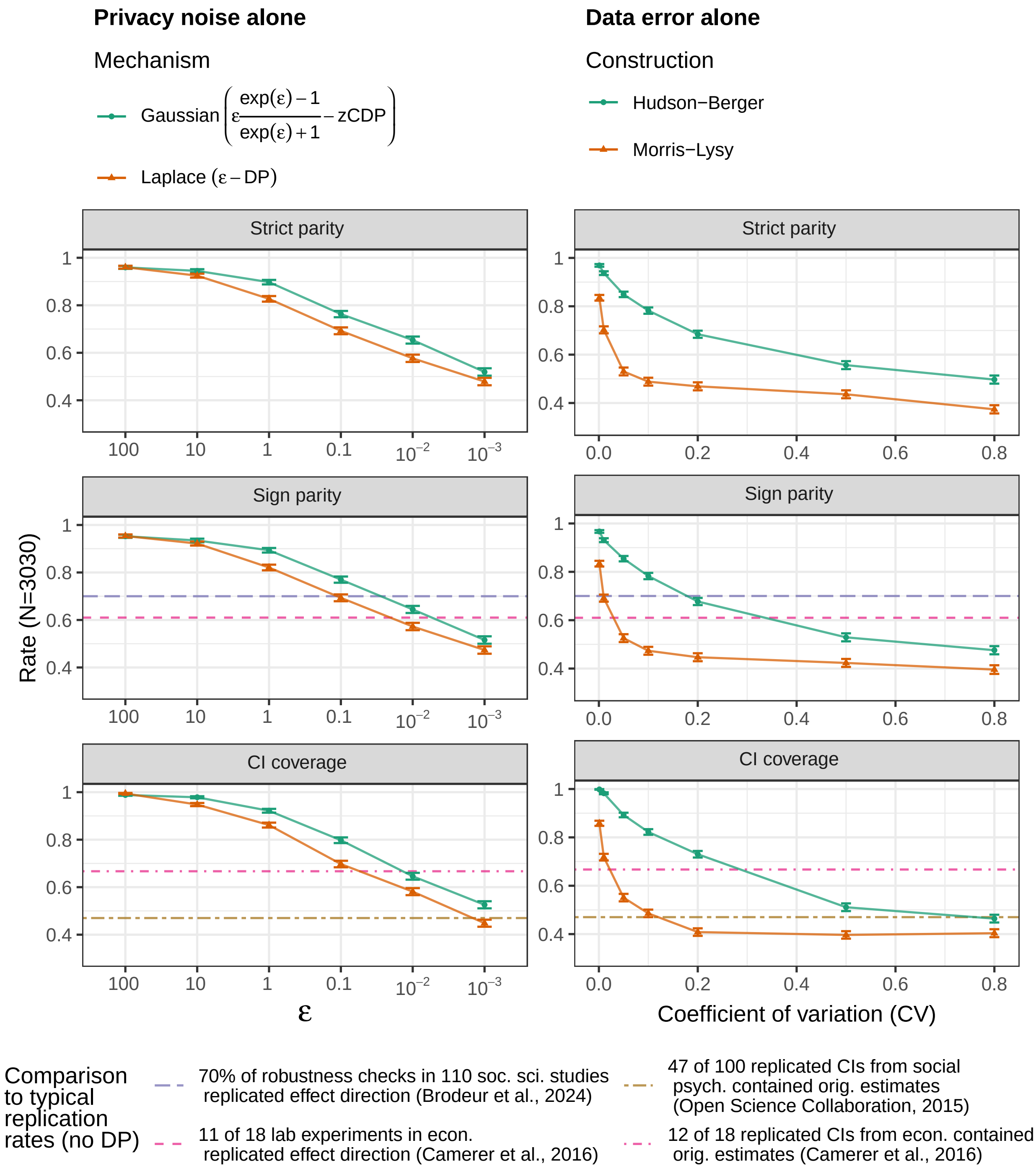


Figure 1: Impacts of DP mechanisms alone (left) and simulated additive data error alone (right) evaluated at confidence level $\alpha = 0.05$, averaged over 303 results times 10 simulations. Error bars depict 90% confidence intervals. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$.

At $\epsilon = 10$ with the $\epsilon\frac{\exp(\epsilon)-1}{\exp(\epsilon)+1}$-zCDP mechanism (a relatively weak privacy guarantee), we observe strict epistemic parity at confidence level $\alpha = 0.1$ in around 95.8% of simulated findings (over 10 replicates per study)—simulations where the results, when reproduced using differentially private data, still support the authors' original claims. The rate of parity is 90.9% at $\epsilon = 1$ (close to common industry settings) and 77.4% at $\epsilon = 0.1$ (a relatively strong guarantee). We observe similar rates of sign parity (matching sign and significance), a metric not based on our interpretation of authors' claims. Almost all of the disparities we observe are originally significant results nullified by noise. The rate of originally *insignificant* results found to be *significant* after noise is very small (1.8% at $\epsilon = 0.1$). Using the ϵ-pure DP Laplace mechanism (the strongest kind of DP guarantee) noticeably decreases the rate of strict parity, particularly for stronger privacy budgets ($\epsilon < 10$)—by 5.8% percentage points at $\epsilon = 1$, for example.

These levels of disparity are lower than the rates observed in large-scale traditional replication studies in economics and other fields. Replicating 110 studies from economics and political science, Brodeur *et al.* [5] find that only 70% of robustness checks (different weights, control variables, estimation methods, fixed effects, etc.) recover a significant effect in the same direction ($p < 0.05$). In our experiments with noise injection, this rate of sign parity occurred at stronger levels of privacy between $\epsilon = 0.1$ and $\epsilon = 0.01$ (Gaussian mechanism).[5] Similarly, a replication of 18 laboratory experiments from economics [3] found a significant effect in the same direction ($p < 0.05$) for only 11 studies (61% sign parity)—equivalent to the effects of the Gaussian mechanism with budget around $\epsilon = 0.01$ in our experiments. The OpenScience project in social psychology found that only 47 of 100 replicated 95% confidence intervals contained their original effect sizes [4]; in our experiments, this rate of CI coverage (47%) occurred only when the privacy guarantee is very strong ($\epsilon < 0.01$ with the Gaussian mechanism).

Are these levels of disparity acceptable to researchers? In a survey of 1,028 American Economic Association members [53], around 60% of economics researchers said they would be willing to accept estimates changing significance levels (from $p < 0.05$) at a rate of up to 10% before they would sacrifice access to noisy estimates based on administrative data.[6] In our sample, sign disparity (similar to significance mismatch)[7] at $\alpha = 0.05$ reaches 10.7% at $\epsilon = 1$ with the zCDP Gaussian

[5] Moreover, different analysts may disagree at similar rates when replicating with the *same* dataset (same data, different team, different methods). When given the same dataset and research question, 20 of 29 analysts found a significant positive effect while 9 did not observe a significant relationship [52]—equivalent to 69% sign parity, similar to the rate of successful robustness checks [5].

[6] These self-reported preferences may change with additional information and may not match revealed preferences in specific contexts. Notably, over 55.2% of respondents said they had "Never heard of the concept" of differential privacy and an additional 24.8% said they had heard of the term but were "not familiar with any of the details" [53].

[7] Williams *et al.* [53] define significance mismatch as "the relative frequency with which a noisy estimate has a different statistical significance (assume 0.05 level) than the estimate without noise" (Appendix B.3). Respondents may have interpreted significance mismatch as sign disparity at $\alpha = 0.05$ (which would count as mismatches cases where

mechanism. Around 60% of economics researchers said they were willing to accept estimates switching sign (similar to sign reversal)[8] at a rate of up to 5% before they would sacrifice access to noisy estimates based on administrative data. We find that this occurs less often; even with a very strong budget $\epsilon = 0.01$, significant positive effects flip to significant negative effects, or vice versa, at a rate of only around 1.4% ($\alpha = 0.05$; SM Figure C.3).

4 Antecedents of epistemic disparity

Table 2 summarizes a linear model of the effect of various mechanism, data, and result characteristics on epistemic parity. We regress over the average strict epistemic parity at $\alpha = 0.1$ for each result ($N = 303$) over all 10 simulations, controlling for study fixed effects. We include all experimental conditions: for privacy (both mechanisms), 50 values of ϵ spaced evenly on a log scale in $[10^{-5}, 10^{3}]$.

Data characteristics. Column 1 presents the impacts of the basic privacy mechanism. Each statistic in the datasets in our sample is the product of one or more statistical queries (noisy counts, noisy sums, and noisy means). Those queries receive random noise from the Laplace and Gaussian mechanisms based on (a) the privacy budget ϵ or $\rho = \epsilon\frac{\exp(\epsilon)-1}{\exp(\epsilon)+1}$ and (b) the global sensitivity of the query (Def. 2). SM Table C.1 demonstrates this model in isolation; the ratio of the root mean squared deviation applied to each component query to the original range of the query is almost completely explained by $\log \epsilon$, the ratio of sensitivity to the original range of the statistic, and the choice of mechanism. The total noise added to each statistic depends on the noise added to its component queries and on the operation used to combine them (e.g., a mean reconstructed with noisy sum over noisy count); so in Column 1, we also control for the type of statistic.

As expected, epistemic parity decreases with sensitivity and increases with ϵ. Within the values of ϵ we tested, a 10-fold increase in ϵ corresponds to a 3.9 percentage point increase in the rate of strict epistemic parity. This estimate is robust to the introduction of other data- and result-related controls. Using zCDP with the Gaussian mechanism corresponds to a $3.4 - 0.2 \log \epsilon$ percentage point increase over the Laplace mechanism (the Gaussian mechanism performs slightly better for $\epsilon < 10$; see Figure 1).

Column 3 adds characteristics of the original regression analysis used to produce the findings.

estimates maintained significance but switched sign), but they may also have interpreted it more literally (a mismatch is any change in significance, up or down, regardless of sign). Under the latter definition, rates of significance mismatch are slightly lower (Figure C.2).

[8] Williams *et al.* [53] define sign mismatch as "the relative frequency with which a noisy estimate is expected to have a different sign (positive or negative) than an estimate without noise" (Appendix B.3). Respondents may have interpreted sign mismatch as sign reversal (excluding initially insignificant results), but they may also have interpreted it more literally (any change in sign, regardless of significance). Under the latter definition, the rate of sign mismatch reaches 4.4% at $\epsilon = 1$ with the zCDP Gaussian mechanism (Figure C.2).

Especially important is the size of the population used to produce each statistic. We proxy for this variable by labeling the time and panel indices for each regression (SM Figures C.9–C.10): compared to regressions over small-population statistics (county-, tract-, block-, school district- , and prefecture-level), regressions over larger-population statistics (1st divisions including states, districts, and provinces) correspond to epistemic parity 9.4 percentage points higher on average. These larger area statistics usually have the same sensitivity as their smaller counterparts (e.g., a simple population count with sensitivity 1), and thus less noise proportional to their size.

Query sensitivity helps explain a few outlier disparities we observe even when very little noise is added (Figure 2a). For example, at $\epsilon = 1000$, an extremely low privacy setting, there are still disparities in a rare 0.8% of findings (Laplace mechanism), all from just 7 results. All of these disparities can be attributed to one or two very sensitive queries over small populations—for example, hospitalizations per million children aged 1 to 5 [49] or monthly neighborhood crime rates [54]. In practice, some of these queries could be further optimized. For example, several are medians, which have particularly large global sensitivity but could be optimized with access to the private microdata and more advanced mechanisms [55]. Removing noise for privacy from only these high-sensitivity queries eliminates these high-ϵ disparities in all cases.

After controlling for study fixed effects, we find no significant effect based on whether the independent (treatment) or dependent (outcome) variables are noised or whether the regression used instrumental variables. However, the privacy-utility curve for findings utilizing IV regressions is noticeably steeper than for other regressions; IV estimates tend to have result in lower rates of epistemic parity (SM Figure C.7).

Epistemic factors. Notably, disparity is impacted by the strength and specificity of the authors' original epistemic claims (Column 2). Findings based on initially weaker evidence are more likely to be contradicted by noisy estimates. Figure 2 depicts this finding graphically with results from the ϵ-DP Laplace mechanism and the standardized effect size metric used by Open Science Collaboration [4] (see §A.2).

This finding mirrors robustness research in social science. Brodeur *et al.* [5], for example, find that half of robustness checks of point estimates significant at $0.05 < p < 0.1$ succeed, compared to 70% of robustness checks of higher-effect point estimates $p < 0.05$.

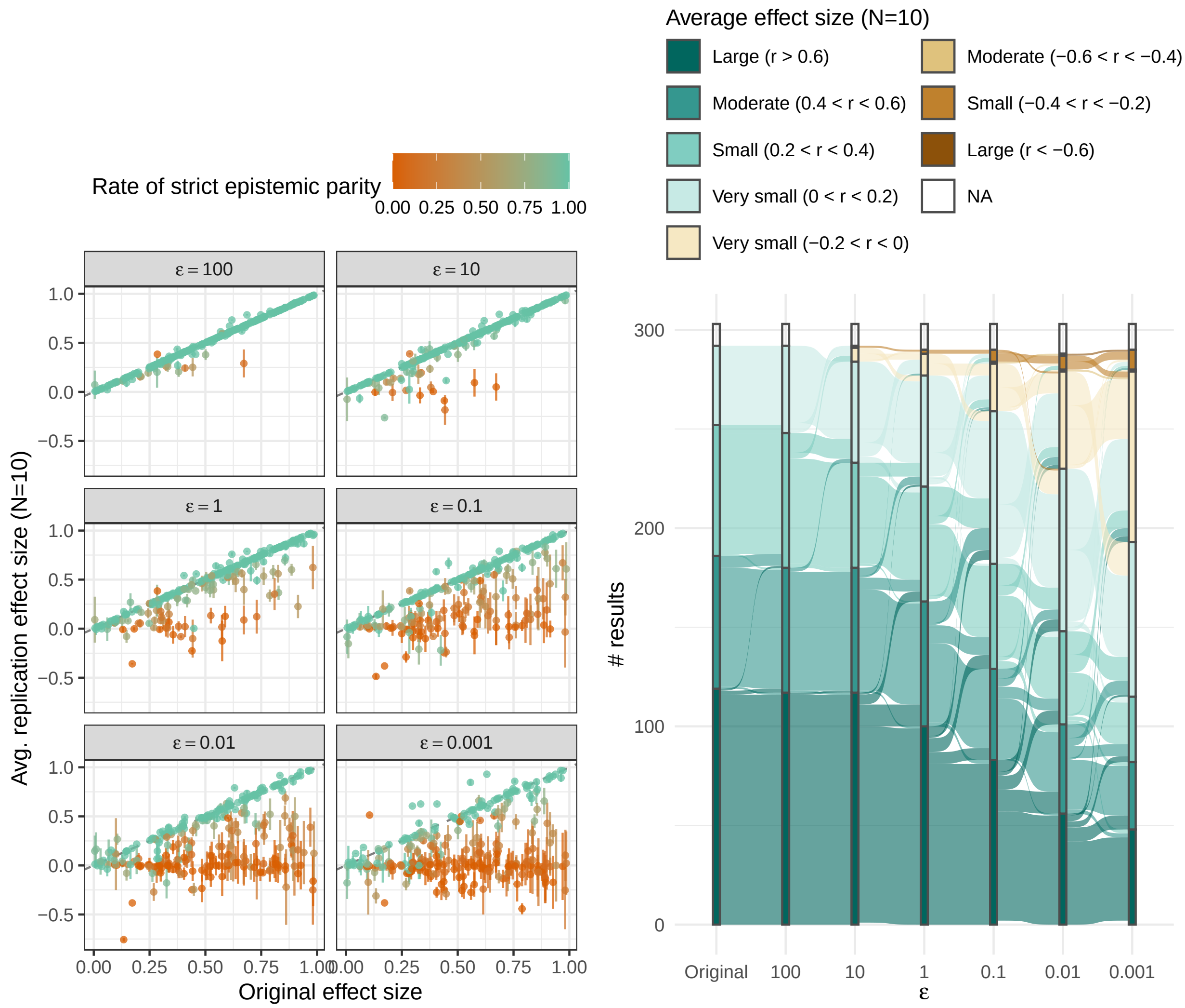
The decrease in effect size is primarily attributable to attenuation, the classic observation that regression estimates deflate and standard errors inflate in the presence of measurement error [56, 57]. Attenuation is clearly visible in the downward shift in the distribution of effect sizes as ϵ decreases (SM Figures C.2, C.4). In particular, adding noise to the outcome variable in a regression leads to 23.0 percentage points lower strict epistemic parity, all else equal. In this ideal case, estimates with additional measurement error are systematically smaller in magnitude, below the original effect size; most of the disparities we document fit this profile (Figure 2a). In robustness checks of 110 social science studies, Brodeur *et al.* [5] observe significance levels 77% of the original; in replication of 18 economics laboratory experiments, Camerer *et al.* [3] similarly observe effect sizes attenuated to 66% of their original value on average.

But in noisier (e.g., lower-powered) settings where the true effect is small, measurement error may increase, not decrease, estimates even as the standard errors increase [58]. Selecting for significance could lead authors to make positive claims about seemingly large estimates when in fact they are drawn from the upper end of the distribution of possible effects. This phenomenon could also help explain why some disparities occur even when little noise is added—the claims are especially sensitive to small amounts of noise despite relatively little change to the effect size (Figure 2a). For $\epsilon > 1$, epistemic disparities come almost exclusively from findings with initially lower effect sizes (e.g., $r < 0.6$; Figure 2a). Most disparities occur when the original claim requires a coefficient with a certain sign or within a certain interval (SM Figure C.6). Null findings (of statistical insignificance), on the other hand correspond to 24.0 percentage points higher epistemic parity than standard direction of effect claims, all else equal. The CI coverage and difference in effect size, which do not depend on the authors' claims about their results, are less sensitive to small amounts of noise. (The rate of 90% CI coverage is 99.7% with the Laplace mechanism at $\epsilon = 1000$.)

Consequences. What do these antecedents mean for different types of research? In general, the preceding analysis suggests that studies or subfields may be particularly impacted if they use more statistics that require privacy protection, rely on statistics about smaller subgroups (e.g., blocks and counties), use more covariates, make claims with more narrow quantitative bounds, or rely on weaker effect sizes. (Studies in Macroeconomics & Monetary Economics have noticeably higher rates of epistemic parity compared to areas such as Labor & Demographic Economics, for example; SM Figure C.11.)

We observe similar trends in the rate of epistemic parity across types of personal data (SM Figure C.12). Regressions using statistics about age, education, and sex/gender tend to have lower rates of epistemic parity, possibly because these variables are often crossed with other variables to create statistics about smaller subgroups. This could have consequences for studies of subgroup differences, an ongoing source of concern for researchers [59]. Election and crime statistics tend to have slightly higher rates of parity, possibly because these statistics in our datasets mostly involve large, simple counts (e.g., of votes or crimes committed).

We also cross-section the privacy utility rates by the type of data source: administrative, census,



(a) Bars depict a 90% confidence interval. Epistemic parity computed at significance level $p < 0.1$.

(b) Horizontal bars depict the number of results whose average effect size change from one bin to another as ϵ decreases. NA indicates reproductions that failed to complete (e.g., a regression failed to converge).

Figure 2: Change in standardized effect size (r, see §A.2), averaged over 10 simulations, as a result of applying the Laplace mechanism with increasing privacy parameter ϵ. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$.

survey, or blended estimates. We find that rates of epistemic parity are already generally lower for survey products, and generally higher for blended estimates under our assumptions (SM Figure C.13). The trade-off curve for census products is nearly flat for $\epsilon > 1$; otherwise, the curves for most data types are similar to the average. However, the type of data product can have great effects on the noise required for differential privacy which are not accounted for in our mechanisms. For example, in our sensitivity analysis, we assume each person contributes equally to each statistic, but many survey products and other statistical estimations are weighted to give more influence to certain subgroups. And while applying an algorithm to a simple random sample can yield smaller privacy loss than applying the same algorithm to the entire population, this amplification effect may not hold for other survey designs [60–62]. Work is still ongoing to explore these considerations and develop optimal DP-compatible design strategies for surveys [61–63] and small-area estimation [64].

5 Impacts of existing data error

Social statistics contain many different kinds of existing data error [32, 33]. To compare the effects of DP mechanisms to the possible impacts of existing data error, we simulate possible alternative versions of the dataset by drawing counterfactual observations from a normal distribution conditioned on the original statistics. We test two admissable constructions using shrinkage estimation from multi-level Bayesian modeling [65]: the Hudson-Berger construction [66], which generally shrinks estimates more when sampling variance is high, and the Morris-Lysy construction, which shrinks estimates more when sampling variance is low [67].

We find that larger epistemic disparities occur if statistics contain even modest amounts of additive data error. U.S. Census Bureau American Community Survey guidance describes a 1.5% coefficient of variation—the ratio of standard deviation to the estimate—as "very small" and "very reliable" for aggregated count data [68]. Simulating normally distributed data error with a coefficient of variation of just 1% results in sign disparity in 4.7% of counterfactual findings (Hudson-Berger construction). A 5% coefficient of variation increases sign disparity to 12.7%. Under the Morris-Lysy construction, disparities are much, much higher—even just a coefficient of variation of 0.1% results in sign disparities in 14.0% of findings.

These choices of the coefficient of variation may be conservative. Prior to 2022, Census Bureau statistical quality standard F1-6 required that the majority of key estimates have a coefficient of variation less than 30% [69]; more complicated data products, such as the school district-level Small Area Income and Poverty Estimates, are published with coefficients as high as 67% [70].

Moreover, researchers may well be overconfident about the robustness of existing research. Asked to predict the robustness of 17 non-experimental *American Economic Review* papers, 359 economists overestimated their robustness reproducibility (46% sign parity on average at $p < 0.05$) by about 15 percentage points [71]. Similar overconfidence may extend to robustness to existing data error.

Given the potentially large effects of additive error on epistemic disparity in research based on counterfactual productions of these social statistics, it is crucial to also evaluate privacy mechanisms *on the margin* of these existing errors. Figure 3 presents the marginal impact of adding the $\epsilon\frac{\exp(\epsilon)-1}{\exp(\epsilon)+1}$-zCDP Gaussian mechanism after simulating data error with the Hudson-Berger construction. With a small amount of data error (coeff. of variation 0.1%), a budget of $\epsilon = 1$ increases epistemic disparity by only 7.6 percentage points in addition to the base effects of data error—less than the 9.1% rate of disparity estimated without accounting for data error (zCDP Gaussian mechanism). Moreover, the marginal impact decreases as the magnitude of simulated data error increases: when the coefficient of variation is 20%, still below the Census Bureau's 2022 quality standard, the privacy mechanism increases disparity by only 3.4 percentage points.

6 Discussion

Researchers and data curators concerned about noise infusion for privacy may take some encouragement from these results. At privacy budgets typical in industry, the chance that privacy noise nullifies or overturns findings is much less than observed in typical robustness checks and replications [3, 5]. And claims with strong statistical support are less likely to be affected.

Yet the problem of noise infusion for privacy forces us to confront deeper issues with robustness in social science. Modest amounts of noise (from either differential privacy mechanisms *or* existing data error) can have noticeable impacts on the scientific findings in our sample. Should authors continue to use the same procedures, as we imagine here, the potential impacts of existing data error, much less widespread adoption of differential privacy, appear non-trivial. Numerous studies have shown that replications of results from psychology and economics do not support the original conclusions at similar or higher rates than what we observe [3, 4, 40, 52]. Other studies similarly show how findings can be sensitive to small changes in the data; Broderick *et al.* [72], for example, find that the results of some economics papers can be overturned by removing less than 1% of the sample.

Acknowledging and accounting for the impacts of existing data error, however, clarifies the trade-offs associated with privacy protections: we find that the relative impacts of differential privacy mechanisms are noticeably slimmer after accounting for the effects of existing data error.

Our methods simulate a scenario with widespread, inefficient DP adoption; the extent to which our results translate to real-world deployments depends on the manner of adoption. The privacy mechanisms tested in this analysis are simple and likely suboptimal. With more effort and expertise, their design could be optimized to the individual context of each data product, reducing disparities. Moreover, our setup assumes near-total adoption. But not all statistics in each study are likely to be made private, not all organizations may implement differential privacy for every statistic, and some statistics may remain invariant for operational or legal reasons, as in the 2020 Decennial Census [7], further reducing disparities. There are still some ways our results may underestimate the impact

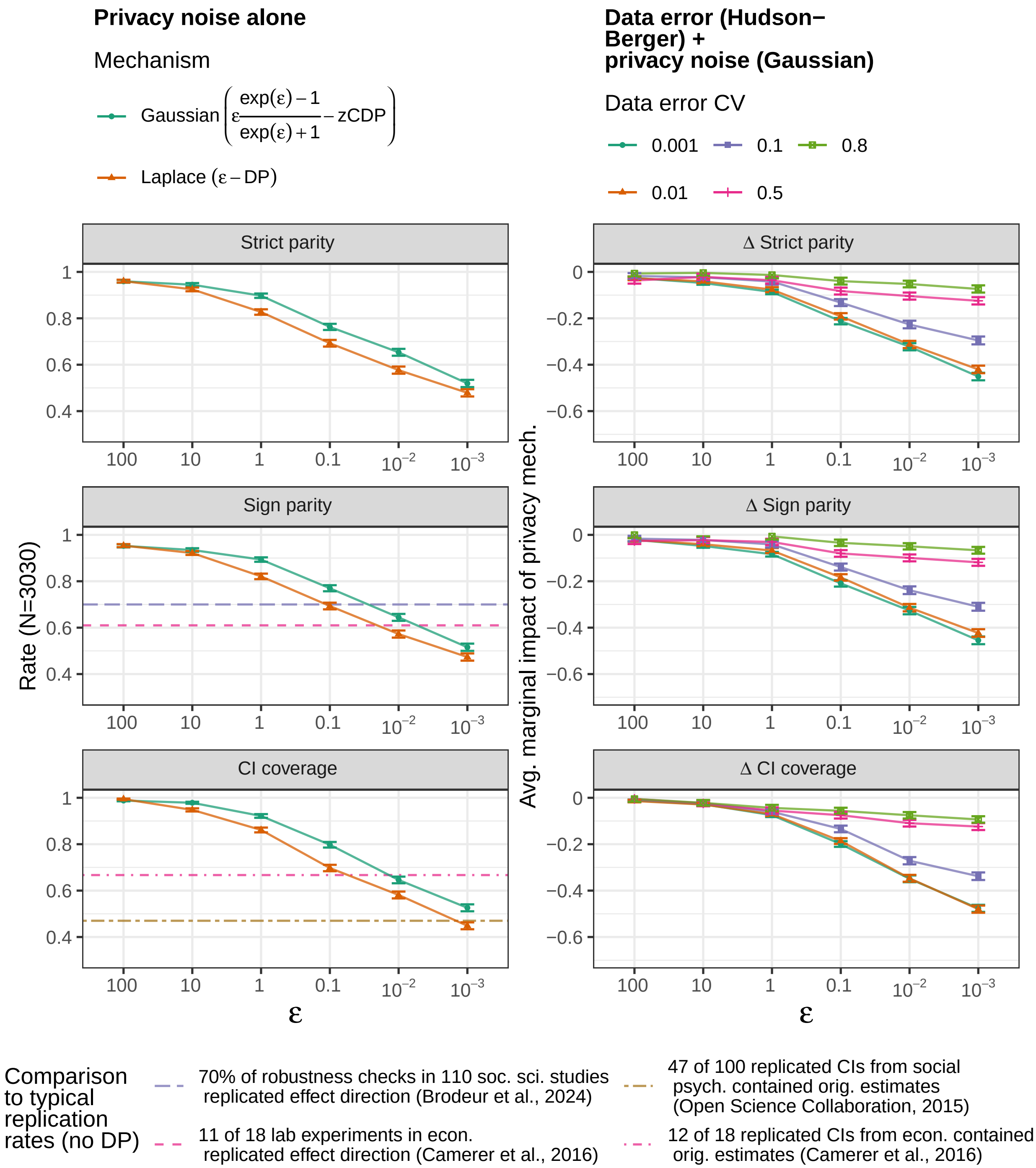


Figure 3: Impacts of DP mechanisms alone (left) and the marginal impact of the zCDP Gaussian mechanism *after* various levels of data error (right), evaluated at confidence level $\alpha = 0.05$, averaged over all results and 10 simulations. Error bars depict 90% confidence intervals. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$.

of differential privacy: in the real world, these statistics have to share privacy budget with other statistics within large data products, and for sensitive survey, blended, and estimated statistics, the noise required to achieve DP may be higher than we assume here. Future work may build on these empirical findings to explore optimal mechanism designs and privacy budget allocations tailored for social science research.

More importantly, this framing challenges the assumption that differential privacy poses an insurmountable obstacle to science. Social science data already includes measurement error, missing values, and other distortions. Differential privacy may be viewed as only a modern addition to these existing sources of data error [32, 33, 73]—an addition meant to protect individuals' privacy and, perhaps, open new opportunities for research on previously inaccessible private dataset.

As our results reiterate, statistical significance is a less useful signal in noisier settings and may result in exaggerated estimates of effect size [58]. There are long-established methods for calibrating regression estimates to account for additive error [57, 74]. Ideally, authors could "robustify" their claims with noise-aware data processing procedures. Recent work proposes methods for multiple imputation and data cleaning for differentially private data specifically [75, 76]. Agarwal & Singh [77], for example, propose a method that could recover the main results of Autor *et al.* [78] (also included in our sample) with equivalent precision down to $\epsilon = 0.01$. There are already examples of this kind of adaptation in practice: authors developed new methods to account for injected noise when Meta released its differentially private Facebook URLs Dataset [76, 79]. However, these efforts are nascent and not widely known; indeed, the majority of economists are not even familiar with differential privacy [53].

While adopting new methods is costly, there can be benefits to scientific practice. For example, Dwork *et al.* [80] and Echenique & He [81] show that by enforcing provable stability in the outputs of statistical analyses, differentially private methods can also inhibit p-hacking and other dubious scientific techniques that rely on overfitting and adaptive data slicing. Indeed, we show that weaker initial effects and overly precise claims are less likely to replicate under differential privacy. Exploring these solutions may call into question the sharp privacy-utility trade-off commonly assumed in policy debates and point the way towards practices that are both more private and more robust [82].

Acknowledgments

We sincerely thank our research assistants Annie Qian and Donna Zhu for their extensive contributions to data collection and replication, and we thank our research assistant Xingyu Chen for his contributions to our replication software. Thanks also to Zeid El-Kilani, Priyanka Nanayakkara, Roy Rinberg, Lucas Rosenblatt, and Steven Wu for feedback on earlier drafts of this work.

Funding. This material is based upon work supported by the National Science Foundation under Grant No. 2319919.

References

1. Hotz, V. J. *et al.* Balancing Data Privacy and Usability in the Federal Statistical System. *Proceedings of the National Academy of Sciences* **119,** e2104906119. https://www.pnas.org/doi/abs/10.1073/pnas.2104906119 (Aug. 2, 2022).
2. Ruggles, S., Fitch, C., Magnuson, D. & Schroeder, J. Differential Privacy and Census Data: Implications for Social and Economic Research. *403 AEA Papers and Proceedings* **109,** 403–408. https://doi.org/10.1257/pandp.20191107 (2019).
3. Camerer, C. F. *et al.* Evaluating Replicability of Laboratory Experiments in Economics. *Science* **351,** 1433–1436. https://www.science.org/doi/10.1126/science.aaf0918 (Mar. 25, 2016).
4. Open Science Collaboration. Estimating the Reproducibility of Psychological Science. *Science* **349,** aac4716. https://www.science.org/doi/10.1126/science.aac4716 (Aug. 28, 2015).
5. Brodeur, A. *et al. Mass Reproducibility and Replicability: A New Hope* Working Paper 107 (I4R Discussion Paper Series, 2024). https://www.econstor.eu/handle/10419/289437.
6. King, G. & Persily, N. *Unprecedented Facebook URLs Dataset Now Available for Academic Research through Social Science One* Social Science One. https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one.
7. Abowd, J. M. *et al.* The 2020 Census Disclosure Avoidance System TopDown Algorithm. *Harvard Data Science Review*. https://hdsr.mitpress.mit.edu/pub/7evz361i/release/1 (Special Issue 2 June 24, 2022).
8. Abowd, J. M. & Hawes, M. B. Confidentiality Protection in the 2020 US Census of Population and Housing. *Annual Review of Statistics and Its Application* **10,** 119–144. http://arxiv.org/abs/2206.03524 (Mar. 10, 2023).
9. Narayanan, A. & Shmatikov, V. Myths and Fallacies of "Personally Identifiable Information". *Communications of the ACM* **53,** 24–26. https://dl.acm.org/doi/10.1145/1743546.1743558 (June 2010).
10. Sweeney, L. K-Anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **10,** 557–570. https://doi.org/10.1142/S0218488502001648 (Oct. 1, 2002).
11. Sweeney, L., Abu, A. & Winn, J. *Identifying Participants in the Personal Genome Project by Name (A Re-identification Experiment)* http://arxiv.org/abs/1304.7605. Pre-published.
12. Narayanan, A. & Shmatikov, V. *How To Break Anonymity of the Netflix Prize Dataset* http://arxiv.org/abs/cs/0610105. Pre-published.

13. Cohen, A. & Nissim, K. Linear Program Reconstruction in Practice. *Journal of Privacy and Confidentiality* **10.** https://journalprivacyconfidentiality.org/index.php/jpc/article/view/711 (1 Jan. 14, 2020).

14. Abowd, J. M. *et al. The 2010 Census Confidentiality Protections Failed, Here's How and Why* http://arxiv.org/abs/2312.11283. Pre-published.

15. Dick, T. *et al.* Confidence-Ranked Reconstruction of Census Microdata from Published Statistics. *Proceedings of the National Academy of Sciences* **120,** e2218605120. https://www.pnas.org/doi/10.1073/pnas.2218605120 (Feb. 21, 2023).

16. Steed, R., Qing, D. & Wu, S. Quantifying Privacy Risks of Public Statistics to Residents of Subsidized Housing. *Harvard Data Science Review.* https://hdsr.mitpress.mit.edu/pub/eb9hacc7/release/2 (Special Issue 6 Aug. 20, 2025).

17. Flaxman, A. & Keyes, O. The Risk of Linked Census Data to Transgender Youth: A Simulation Study. *Journal of Privacy and Confidentiality* **15.** https://journalprivacyconfidentialit org/index.php/jpc/article/view/891 (1 Mar. 31, 2025).

18. boyd, d. & Sarathy, J. Differential Perspectives: Epistemic Disconnects Surrounding the U.S. Census Bureau's Use of Differential Privacy. *Harvard Data Science Review.* https://hdsr.mitpress.mit.edu/pub/3vj5j6i0/release/1 (Special Issue 2 June 24, 2022).

19. Dinur, I. & Nissim, K. *Revealing Information While Preserving Privacy* in *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems - PODS '03* The Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium (ACM Press, San Diego, California, 2003), 202–210. http://portal.acm.org/citation.cfm?doid=773153.773173.

20. Dwork, C., McSherry, F., Nissim, K. & Smith, A. *Calibrating Noise to Sensitivity in Private Data Analysis* in *Proceedings of the Third Conference on Theory of Cryptography* (Springer-Verlag, Berlin, Heidelberg, Mar. 4, 2006), 265–284. https://doi.org/10.1007/11681878_14.

21. Adeleye, T. *et al. Publishing Wikipedia Usage Data with Strong Privacy Guarantees* http://arxiv.org/abs/2308.16298. Pre-published.

22. Desfontaines, D. *A List of Real-World Uses of Differential Privacy* Ted is writing things. https://desfontain.es/privacy/real-world-differential-privacy.html.

23. Nanayakkara, P., Ghazi, E. & Vadhan, S. *Practitioners' Perspectives on a Differential Privacy Deployment Registry* http://arxiv.org/abs/2509.13509. Pre-published.

24. Khavkin, M. & Toch, E. Differential Privacy Configurations in the Real World: A Comparative Analysis. *IEEE Transactions on Knowledge and Data Engineering* **37,** 6334–6351. https://ieeexplore.ieee.org/abstract/document/11143893 (Nov. 2025).

25. Biden, J. R. *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* The White House, Oct. 30, 2023. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

26. National Science and Technology Council. *National Strategy to Advance Privacy-Preserving Data Sharing and Analytics* (Executive Office of the President, Mar. 2023).

27. Leanor, M. McEachin Casts Doubt on Census Bureau's Plans for Population Count Data. *Richmond Times-Dispatch.* https://richmond.com/news/state-and-regional/govt-and-politics/mceachin-casts-doubt-on-census-bureaus-plans-for-population-count-data/article_78b9dd06-c21d-5acc-b564-00fbc5934607.html (Sept. 21, 2020).

28. Ochoa, S. A. & Minnis, T. A. *Preliminary Report: Impact of Differential Privacy & the 2020 Census on Latinos, Asian Americans and Redistricting* Report (MALDEF & Asian Americans Advancing Justice, Apr. 2021). https://www.maldef.org/wp-content/uploads/2021/04/FINAL-MALDEF-AAJC-Differential-Privacy-Preliminary-Report-4.5.2021-1.pdf.

29. Oberski, D. & Kreuter, F. Differential Privacy and Social Science: An Urgent Puzzle. *Harvard Data Science Review* **2.** https://hdsr.mitpress.mit.edu/pub/g9o4z8au/release/8 (Jan. 31, 2020).

30. Childs, J. H. & Bureau, U. S. C. Trust and Credibility in the U.S. Federal Statistical System. *Survey Methods: Insights from the Field (SMIF).* https://surveyinsights.org/?p=10663 (Feb. 22, 2019).

31. Rosenblatt, L. *et al. Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy* http://arxiv.org/abs/2208.12700. Pre-published.

32. Groves, R. M. & Lyberg, L. Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly* **74,** 849–879. https://doi.org/10.1093/poq/nfq065 (Jan. 1, 2010).

33. Steed, R., Liu, T., Wu, Z. S. & Acquisti, A. Policy Impacts of Statistical Uncertainty and Privacy. *Science* **377,** 928–931. https://www.science.org/doi/10.1126/science.abq4481 (Aug. 26, 2022).

34. Abowd, J. M. & Schmutte, I. M. An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices. *American Economic Review* **109,** 171–202 (2019).

35. Kenny, C. T. *et al.* The Use of Differential Privacy for Census Data and Its Impact on Redistricting: The Case of the 2020 U.S. Census. *Science Advances* **7,** eabk3283. https://www.science.org/doi/abs/10.1126/sciadv.abk3283 (2021).

36. Pujol, D. *et al. Fair Decision Making Using Privacy-Protected Data* in *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, Inc, Jan. 2020), 189–199.

37. Brummet, Q., Mulrow, E. & Wolter, K. The Effect of Differentially Private Noise Injection on Sampling Efficiency and Funding Allocations: Evidence From the 1940 Census. *Harvard Data Science Review*. https://hdsr.mitpress.mit.edu/pub/ft9fhmku/release/1 (Special Issue 2 June 24, 2022).
38. Barrientos, A. F., Williams, A. R., Snoke, J. & Bowen, C. M. A Feasibility Study of Differentially Private Summary Statistics and Regression Analyses with Evaluations on Administrative and Survey Data. *Journal of the American Statistical Association* **0,** 1–24. https://doi.org/10.1080/01621459.2023.2270795 (ja 2023).
39. Williams, A. R., Barrientos, A. F., Snoke, J. & Bowen, C. M. *Benchmarking DP Linear Regression Methods for Statistical Inference* 2024.
40. Klein, R. A. *et al.* Investigating Variation in Replicability. *Social Psychology* **45,** 142–152. https://econtent.hogrefe.com/doi/10.1027/1864-9335/a000178 (May 2014).
41. IPUMS. *Bibliography* IPUMS. https://bibliography.ipums.org/citations/results?search_terms%5Bmax_year_published%5D=2024&search_terms%5Bmin_year_published%5D=1960.
42. Penberthy, L. SEER Marks 50 Years of Turning Cancer Data into Discovery. *NIH Record* **LXXV,** 12. https://nihrecord.nih.gov/2023/02/03/seer-marks-50-years-turning-cancer-data-discovery (Feb. 3, 2023).
43. Acquisti, A. & Steed, R. Learning to Live with Privacy-Preserving Analytics. *Communications of the ACM* **66,** 24–27. https://dl.acm.org/doi/10.1145/3597173 (June 22, 2023).
44. Inter-University Consortium for Political and Social Research. *Find Data* ICPSR. https://www.icpsr.umich.edu/web/pages/ICPSR/index.html.
45. Kranz, S. *Find Economic Articles with Data* https://ejd.econ.mathematik.uni-ulm.de/.
46. Dwork, C. & Rothblum, G. N. *Concentrated Differential Privacy* http://arxiv.org/abs/1603.01887. Pre-published.
47. Bun, M. & Steinke, T. *Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds* http://arxiv.org/abs/1605.02065. Pre-published.
48. Cohen, K. B. *et al. Three Dimensions of Reproducibility in Natural Language Processing* in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* LREC 2018 (eds Calzolari, N. *et al.*) (European Language Resources Association (ELRA), Miyazaki, Japan, May 2018). https://aclanthology.org/L18-1025.
49. Guidetti, B., Pereda, P. & Severnini, E. “Placebo Tests” for the Impacts of Air Pollution on Health: The Challenge of Limited Health Care Infrastructure. *AEA Papers and Proceedings* **111,** 371–375. https://www.jstor.org/stable/27042333 (2021).
50. Dwork, C., Kohli, N. & Mulligan, D. Differential Privacy in Practice: Expose Your Epsilons! *Journal of Privacy and Confidentiality* **9.** https://journalprivacyconfidentiality.org/index.php/jpc/article/view/689 (2 Oct. 20, 2019).

51. US Census Bureau. *Privacy-Loss Budget Allocation 2023-04-03* Mar. 4, 2023. https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/04-Demonstration_Data_Products_Suite/2023-04-03/2023-04-03_Privacy-Loss_Budget_Allocations.pdf.
52. Silberzahn, R. *et al.* Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices in Psychological Science* **1,** 337–356. https://doi.org/10.1177/2515245917747646 (Sept. 1, 2018).
53. Williams, A. R., Snoke, J., Bowen, C. M. & Barrientos, A. F. Disclosing Economists' Privacy Perspectives: A Survey of American Economic Association Members' Views on Differential Privacy and the Usability of Noise-Infused Data. *Harvard Data Science Review.* https://hdsr.mitpress.mit.edu/pub/n4db7079/release/1 (Aug. 21, 2024).
54. Bellégo, C. & Drouard, J. Fighting Crime in Lawless Areas: Evidence from Slums in Rio de Janeiro. *American Economic Journal: Economic Policy* **16,** 124–159. https://www.aeaweb.org/articles?id=10.1257/pol.20200587 (Feb. 2024).
55. Cummings, R. *et al.* Advancing Differential Privacy: Where We Are Now and Future Directions for Real-World Deployment. *Harvard Data Science Review.* https://hdsr.mitpress.mit.edu/pub/sl9we8gh/release/1 (Jan. 16, 2024).
56. Hausman, J. Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *Journal of Economic Perspectives* **15,** 57–67. https://www.aeaweb.org/articles?id=10.1257/jep.15.4.57 (Dec. 2001).
57. Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition* 2nd ed. 488 pp. (Chapman and Hall/CRC, New York, June 20, 2006).
58. Loken, E. & Gelman, A. Measurement Error and the Replication Crisis. *Science* **355,** 584–585. https://www.science.org/doi/10.1126/science.aal3618 (Feb. 10, 2017).
59. Santos-Lozada, A. R., Howard, J. T. & Verdery, A. M. How Differential Privacy Will Affect Our Understanding of Health Disparities in the United States. *Proceedings of the National Academy of Sciences of the United States of America* **117,** 13405–13412. www.pnas.org/cgi/doi/10.1073/pnas.2003714117 (June 2020).
60. Bun, M., Drechsler, J., Gaboardi, M., McMillan, A. & Sarathy, J. *Controlling Privacy Loss in Sampling Schemes: An Analysis of Stratified and Cluster Sampling* http://arxiv.org/abs/2007.12674. Pre-published.
61. Drechsler, J. & Bailie, J. *The Complexities of Differential Privacy for Survey Data* http://arxiv.org/abs/2408.07006. Pre-published.
62. Seeman, J., Si, Y. & Reiter, J. P. *Differentially Private Finite Population Estimation via Survey Weight Regularization* http://arxiv.org/abs/2411.04236. Pre-published.

63. U.S. Census Bureau. *Disclosure Avoidance Protections for the American Community Survey* Census Blogs. https://www.census.gov/newsroom/blogs/random-samplings/2022/12/disclosure-avoidance-protections-acs.html.
64. Seeman, J., Slavkovic, A. & Reimherr, M. *Private Posterior Inference Consistent with Public Information: A Case Study in Small Area Estimation from Synthetic Census Data* in *Privacy in Statistical Databases* (eds Domingo-Ferrer, J. & Muralidhar, K.) (Springer International Publishing, Cham, 2020), 323–336.
65. Cui, Y., Gong, R., Hannig, J. & Hoffman, K. Technical Comment on “Policy Impacts of Statistical Uncertainty and Privacy”. *Science* **380,** eadf9724. https://www.science.org/doi/10.1126/science.adf9724 (June 2, 2023).
66. Hudson, H. M. *Empirical Bayes Estimation* (1974). https://purl.stanford.edu/cv052xn0322.
67. Morris, C. N. & Lysy, M. Shrinkage Estimation in Multilevel Normal Models. *Statistical Science* **27,** 115–134. https://projecteuclid.org/journals/statistical-science/volume-27/issue-1/Shrinkage-Estimation-in-Multilevel-Normal-Models/10.1214/11-STS363.full (Feb. 2012).
68. U.S. Census Bureau. in *ACS General Handbook* (U.S. Government Printing Office, Washington, D.C., July 2018). https://www.census.gov/content/dam/Census/library/publications/2018/acs/acs_general_handbook_2018.pdf.
69. U.S. Census Bureau. *U.S. Census Bureau Statistical Quality Standards* July 2013. https://www2.census.gov/about/policies/quality/quality-standards-jul2013.pdf.
70. Maples, J. *Calculating Coefficient of Variation for the Minimum Change School District Poverty Estimates and the Assessment of the Impact of Nongeocoded Tax Returns* Statistics 2008–10 (U.S. Census Bureau, 2008), 17.
71. Campbell, D. *et al. The Robustness Reproducibility of the American Economic Review* Working Paper 124 (I4R Discussion Paper Series, 2024). https://www.econstor.eu/handle/10419/295222.
72. Broderick, T., Giordano, R. & Meager, R. *An Automatic Finite-Sample Robustness Metric: When Can Dropping a Little Data Make a Big Difference?* http://arxiv.org/abs/2011.14999. Pre-published.
73. Gong, R. Transparent Privacy Is Principled Privacy. *Harvard Data Science Review.* https://hdsr.mitpress.mit.edu/pub/ld4smnnf/release/5 (Special Issue 2 June 24, 2022).
74. Cook, J. R. & Stefanski, L. A. Simulation-Extrapolation Estimation in Parametric Measurement Error Models. *Journal of the American Statistical Association* **89,** 1314–1328. https://www.jstor.org/stable/2290994 (1994).
75. Blackwell, M., Honaker, J. & King, G. A Unified Approach to Measurement Error and Missing Data: Overview and Applications. *Sociological Methods & Research* **46,** 303–341. https://doi.org/10.1177/0049124115585360 (Aug. 1, 2017).

76. Evans, G. & King, G. Statistically Valid Inferences from Differentially Private Data Releases, with Application to the Facebook URLs Dataset. *Political Analysis* **31,** 1–21. https://www.cambridge.org/core/journals/political-analysis/article/statistically-valid-inferences-from-differentially-private-data-releases-with-application-to-the-facebook-urls-dataset/EF48830D31BCA66C1E6FD5C995F33B71 (Jan. 2023).

77. Agarwal, A. & Singh, R. *Causal Inference with Corrupted Data: Measurement Error, Missing Values, Discretization, and Differential Privacy* http://arxiv.org/abs/2107.02780. Pre-published.

78. Autor, D. H., Dorn, D. & Hanson, G. H. The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review* **103,** 2121–2168. https://www.aeaweb.org/articles?id=10.1257/aer.103.6.2121 (Oct. 2013).

79. Buntain, C., Bonneau, R., Nagler, J. & Tucker, J. A. *Measuring the Ideology of Audiences for Web Links and Domains Using Differentially Private Engagement Data* https://papers.ssrn.com/abstract=3765240. Pre-published.

80. Dwork, C. *et al.* The Reusable Holdout: Preserving Validity in Adaptive Data Analysis. *Science* **349,** 636–638. https://www.science.org/doi/10.1126/science.aaa9375 (Aug. 7, 2015).

81. Echenique, F. & He, K. *Screening p-Hackers: Dissemination Noise as Bait* http://arxiv.org/abs/2103.09164. Pre-published.

82. Martin, K. *Privacy Myths and Mistakes: Paradoxes, Tradeoffs, and the Omnipotent Consumer* 2024. https://kirstenmartin.net/wp-content/uploads/2024/11/Privacy-Myths-DRAFT-10-25.pdf.

83. Barba, L. A. *Terminologies for Reproducible Research* http://arxiv.org/abs/1802.03311. Pre-published.

84. Broman, K. *et al. Recommendations to Funding Agencies for Supporting Reproducible Research* (American Statistical Association, Jan. 18, 2017). https://www.amstat.org/asa/files/pdfs/POL-ReproducibleResearchRecommendations.pdf.

85. Abouk, R. & Adams, S. Texting Bans and Fatal Accidents on Roadways: Do They Work? Or Do Drivers Just React to Announcements of Bans? *American Economic Journal: Applied Economics* **5,** 179–199. https://www.aeaweb.org/articles?id=10.1257/app.5.2.179 (Apr. 2013).

86. Acconcia, A., Corsetti, G. & Simonelli, S. Mafia and Public Spending: Evidence on the Fiscal Multiplier from a Quasi-experiment. *American Economic Review* **104,** 2185–2209. https://www.aeaweb.org/articles?id=10.1257/aer.104.7.2185 (July 2014).

87. El Fatmaoui, A. From High School to Higher Education: Is Recreational Marijuana a Consumption Amenity for US College Students? *Economic Inquiry* **62,** 1024–1045. https://onlinelibrary.wiley.com/doi/abs/10.1111/ecin.13225 (2024).

88. Ajzenman, N., Dominguez, P. & Undurraga, R. Immigration, Crime, and Crime (Mis)Perceptions. *American Economic Journal: Applied Economics* **15,** 142–176. https://www.aeaweb.org/articles?id=10.1257/app.20210156 (Oct. 2023).
89. Stashko, A. Do Police Maximize Arrests or Minimize Crime? Evidence from Racial Profiling in U.S. Cities. *Journal of the European Economic Association* **21,** 167–214. https://doi.org/10.1093/jeea/jvac039 (Feb. 1, 2023).
90. Anderson, D. M. In School and Out of Trouble? The Minimum Dropout Age and Juvenile Crime. *The Review of Economics and Statistics* **96,** 318–331. https://www.jstor.org/stable/43554933 (2014).
91. Anderson, S., Francois, P. & Kotwal, A. Clientelism in Indian Villages. *American Economic Review* **105,** 1780–1816. https://www.aeaweb.org/articles?id=10.1257/aer.20130623 (June 2015).
92. Antecol, H., Bedard, K. & Stearns, J. Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies? *American Economic Review* **108,** 2420–2441. https://www.aeaweb.org/articles?id=10.1257/aer.20160613 (Sept. 2018).
93. Bai, L. & Wu, L. Political Conflict and Development Dynamics: Economic Legacies of the Cultural Revolution. *The Journal of Economic History* **83,** 981–1017. https://www.cambridge.org/core/journals/journal-of-economic-history/article/political-conflict-and-development-dynamics-economic-legacies-of-the-cultural-revolution/796180C225D3D9522E4E22C020F4639C (Dec. 2023).
94. Baker, S. R. Effects of Immigrant Legalization on Crime. *American Economic Review* **105,** 210–213. https://www.aeaweb.org/articles?id=10.1257/aer.p20151041 (May 2015).
95. Le Barbanchon, T. & Sauvagnat, J. Electoral Competition, Voter Bias, and Women in Politics. *Journal of the European Economic Association* **20,** 352–394. https://doi.org/10.1093/jeea/jvab028 (Feb. 1, 2022).
96. Baron, E. J. School Spending and Student Outcomes: Evidence from Revenue Limit Elections in Wisconsin. *American Economic Journal: Economic Policy* **14,** 1–39. https://www.aeaweb.org/articles?id=10.1257/pol.20200226 (Feb. 2022).
97. Beach, B. & Jones, D. B. Gridlock: Ethnic Diversity in Government and the Provision of Public Goods. *American Economic Journal: Economic Policy* **9,** 112–136. https://www.aeaweb.org/articles?id=10.1257/pol.20150394 (Feb. 2017).
98. Bell, B., Costa, R. & Machin, S. Why Does Education Reduce Crime? *Journal of Political Economy* **130,** 732–765. https://www.journals.uchicago.edu/doi/10.1086/717895 (Mar. 2022).
99. Bernecker, A., Boyer, P. C. & Gathmann, C. The Role of Electoral Incentives for Policy Innovation: Evidence from the US Welfare Reform. *American Economic Journal: Economic Policy* **13,** 26–57. https://www.aeaweb.org/articles?id=10.1257/pol.20190690 (May 2021).

100. Bohn, S., Freedman, M. & Owens, E. The Criminal Justice Response to Policy Interventions: Evidence from Immigration Reform. *American Economic Review* **105,** 214–219. https://www.aeaweb.org/articles?id=10.1257/aer.p20151042 (May 2015).

101. Bordignon, M., Nannicini, T. & Tabellini, G. Moderating Political Extremism: Single Round versus Runoff Elections under Plurality Rule. *American Economic Review* **106,** 2349–2370. https://www.aeaweb.org/articles?id=10.1257/aer.20131024 (Aug. 2016).

102. Bostwick, V., Fischer, S. & Lang, M. Semesters or Quarters? The Effect of the Academic Calendar on Postsecondary Student Outcomes. *American Economic Journal: Economic Policy* **14,** 40–80. https://www.aeaweb.org/articles?id=10.1257/pol.20190589 (Feb. 2022).

103. Boulware, K. D. & Kuttner, K. N. Labor Market Conditions and Discrimination: Is There a Link? *AEA Papers and Proceedings* **109,** 166–170. https://www.aeaweb.org/articles?id=10.1257/pandp.20191086 (May 2019).

104. Bove, V. & Gavrilova, E. Police Officer on the Frontline or a Soldier? The Effect of Police Militarization on Crime. *American Economic Journal: Economic Policy* **9,** 1–18. https://www.aeaweb.org/articles?id=10.1257/pol.20150478 (Aug. 2017).

105. Brinkman, J., Coen-Pirani, D. & Sieg, H. The Political Economy of Municipal Pension Funding. *American Economic Journal: Macroeconomics* **10,** 215–246. https://www.aeaweb.org/articles?id=10.1257/mac.20160400 (July 2018).

106. Brollo, F., Nannicini, T., Perotti, R. & Tabellini, G. The Political Resource Curse. *American Economic Review* **103,** 1759–1796. https://www.aeaweb.org/articles?id=10.1257/aer.103.5.1759 (Aug. 2013).

107. Buonanno, P. & Raphael, S. Incarceration and Incapacitation: Evidence from the 2006 Italian Collective Pardon. *American Economic Review* **103,** 2437–2465. https://www.aeaweb.org/articles?id=10.1257%2Faer.103.6.2437&ref=cold-takes (Oct. 2013).

108. Campante, F. R., Do, Q.-A. & Guimaraes, B. Capital Cities, Conflict, and Misgovernance. *American Economic Journal: Applied Economics* **11,** 298–337. https://www.aeaweb.org/articles?id=10.1257/app.20170111 (July 2019).

109. Cascio, E. U. & Lewis, E. G. Cracks in the Melting Pot: Immigration, School Choice, and Segregation. *American Economic Journal: Economic Policy* **4,** 91–117. https://www.jstor.org/stable/23249922 (2012).

110. Caselli, F. & Michaels, G. Do Oil Windfalls Improve Living Standards? Evidence from Brazil. *American Economic Journal: Applied Economics* **5,** 208–238. https://www.aeaweb.org/articles?id=10.1257/app.5.1.208 (Jan. 2013).

111. Chalfin, A. The Long-Run Effect of Mexican Immigration on Crime in US Cities: Evidence from Variation in Mexican Fertility Rates. *American Economic Review* **105,** 220–225. https://www.aeaweb.org/articles?id=10.1257/aer.p20151043 (May 2015).

112. Charles, K. K., Hurst, E. & Notowidigdo, M. J. Housing Booms and Busts, Labor Market Opportunities, and College Attendance. *American Economic Review* **108,** 2947–2994. https://www.aeaweb.org/articles?id=10.1257/aer.20151604 (Oct. 2018).

113. Charles, K. K. & Stephens, M. Employment, Wages, and Voter Turnout. *American Economic Journal: Applied Economics* **5,** 111–143. https://www.jstor.org/stable/43189455 (2013).

114. Chodorow-Reich, G., Feiveson, L., Liscow, Z. & Woolston, W. G. Does State Fiscal Relief during Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act. *American Economic Journal: Economic Policy* **4,** 118–145. https://www.aeaweb.org/articles?id=10.1257/pol.4.3.118 (Apr. 2012).

115. Chu, J., Fisman, R., Tan, S. & Wang, Y. Hometown Ties and the Quality of Government Monitoring: Evidence from Rotation of Chinese Auditors. *American Economic Journal: Applied Economics* **13,** 176–201. https://www.aeaweb.org/articles?id=10.1257/app.20190516 (July 2021).

116. Combs, A. E. & Foster, J. M. The Effects of Homestead Exemptions for Seniors and Disabled People on School Districts. *AERA Open* **7,** 2332858420988712. https://doi.org/10.1177/2332858420988712 (Jan. 1, 2021).

117. Crane, A. D., Koch, A. & Lin, L. Real Effects of Markets on Politics: Evidence from US Presidential Elections. *American Economic Review: Insights* **6,** 73–88. https://www.aeaweb.org/articles?id=10.1257/aeri.20220240 (Mar. 2024).

118. Cunningham, J. P. & Gillezeau, R. Racial Differences in Police Use of Force: Evidence from the 1960s Civil Disturbances. *AEA Papers and Proceedings* **108,** 217–221. https://www.aeaweb.org/articles?id=10.1257/pandp.20181110 (May 2018).

119. DellaVigna, S., Enikolopov, R., Mironova, V., Petrova, M. & Zhuravskaya, E. Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia. *American Economic Journal: Applied Economics* **6,** 103–132. https://www.aeaweb.org/articles?id=10.1257/app.6.3.103 (July 2014).

120. Dhaliwal, T. K. & Bruno, P. The Rural/Nonrural Divide? K–12 District Spending and Implications of Equity-Based School Funding. *AERA Open* **7,** 2332858420982549. https://doi.org/10.1177/2332858420982549 (Jan. 1, 2021).

121. Dias, M., Rocha, R. & Soares, R. R. Down the River: Glyphosate Use in Agriculture and Birth Outcomes of Surrounding Populations. *The Review of Economic Studies* **90,** 2943–2981. https://doi.org/10.1093/restud/rdad011 (Nov. 1, 2023).

122. Dickstein, M. J., Duggan, M., Orsini, J. & Tebaldi, P. The Impact of Market Size and Composition on Health Insurance Premiums: Evidence from the First Year of the Affordable Care Act. *American Economic Review* **105,** 120–125. https://pubs.aeaweb.org/doi/10.1257/aer.p20151083 (May 1, 2015).

123. Dieterle, S., Bartalotti, O. & Brummet, Q. Revisiting the Effects of Unemployment Insurance Extensions on Unemployment: A Measurement-Error-Corrected Regression Discontinuity Approach. *American Economic Journal: Economic Policy* **12,** 84–114. https://www.aeaweb.org/articles?id=10.1257/pol.20160439 (May 2020).

124. Autor, D. H. & Dorn, D. The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review* **103,** 1553–1597. https://www.aeaweb.org/articles?id=10.1257/aer.103.5.1553 (Aug. 2013).

125. Drago, F., Nannicini, T. & Sobbrio, F. Meet the Press: How Voters and Politicians Respond to Newspaper Entry and Exit. *American Economic Journal: Applied Economics* **6,** 159–188. https://www.aeaweb.org/articles?id=10.1257/app.6.3.159 (July 2014).

126. Duquette, N. J. Inequality and Philanthropy: High-income Giving in the United States 1917–2012. *Explorations in Economic History* **70,** 25–41. https://www.sciencedirect.com/science/article/pii/S0014498318300202 (Oct. 1, 2018).

127. Egger, P. & Koethenbuerger, M. Government Spending and Legislative Organization: Quasi-experimental Evidence from Germany. *American Economic Journal: Applied Economics* **2,** 200–212. https://www.aeaweb.org/articles?id=10.1257/app.2.4.200 (Oct. 2010).

128. Eriksen, M. D. & Ross, A. Housing Vouchers and the Price of Rental Housing. *American Economic Journal: Economic Policy* **7,** 154–176. https://www.aeaweb.org/articles?id=10.1257/pol.20130064 (Aug. 2015).

129. Evans, W. N., Muchnick, K. & Rosenlund, O. Virtual Learning in Kindergarten Through Grade 12 During the COVID-19 Pandemic and Chronic Absenteeism. *JAMA Network Open* **7,** e2429569. https://doi.org/10.1001/jamanetworkopen.2024.29569 (Aug. 21, 2024).

130. Fernandez, J. M. Hurricane Maria and La Crisis Boricua on Health-Care Supply in Puerto Rico. *AEA Papers and Proceedings* **111,** 598–601. https://www.aeaweb.org/articles?id=10.1257/pandp.20211116 (May 2021).

131. Ferraz, C. & Finan, F. Electoral Accountability and Corruption: Evidence from the Audits of Local Governments. *American Economic Review* **101,** 1274–1311. https://pubs.aeaweb.org/doi/10.1257/aer.101.4.1274 (June 1, 2011).

132. Fisman, R. & Wang, Y. The Distortionary Effects of Incentives in Government: Evidence from China's "Death Ceiling" Program. *American Economic Journal: Applied Economics* **9,** 202–218. https://www.aeaweb.org/articles?id=10.1257/app.20160008 (Apr. 2017).

133. Fitzpatrick, M. D. & Lovenheim, M. F. Early Retirement Incentives and Student Achievement. *American Economic Journal: Economic Policy* **6,** 120–154. https://www.aeaweb.org/articles?id=10.1257/pol.6.3.120 (Aug. 2014).

134. Forman, C., Goldfarb, A. & Greenstein, S. The Internet and Local Wages: A Puzzle. *American Economic Review* **102,** 556–575. https://www.aeaweb.org/articles?id=10.1257/aer.102.1.556 (Feb. 2012).

135. Fortin, N. M. Higher-Education Policies and the College Wage Premium: Cross-State Evidence from the 1990s. *American Economic Review* **96,** 959–987. https://www.aeaweb.org/articles?id=10.1257/aer.96.4.959 (Sept. 2006).

136. Furtado, D. & Kong, H. How Do Low-Education Immigrants Adjust to Chinese Import Shocks? Evidence Using English Language Proficiency. *European Economic Review* **163,** 104681. https://www.sciencedirect.com/science/article/pii/S0014292124000102 (Apr. 1, 2024).

137. Ellison, G. & Ellison, S. F. Tax Sensitivity and Home State Preferences in Internet Purchasing. *American Economic Journal: Economic Policy* **1,** 53–71. https://www.aeaweb.org/articles?id=10.1257/pol.1.2.53 (Aug. 2009).

138. González, L. The Effect of a Universal Child Benefit on Conceptions, Abortions, and Early Maternal Labor Supply. *American Economic Journal: Economic Policy* **5,** 160–188. https://www.aeaweb.org/articles?id=10.1257/pol.5.3.160 (Aug. 2013).

139. Gonzalez, R., Jabri, R. & Komisarow, S. When Does Crime Respond to Punishment?: Evidence from Drug-Free School Zones. *Journal of Urban Economics* **147,** 103763. https://www.sciencedirect.com/science/article/pii/S0094119025000282 (May 1, 2025).

140. Grier, R. The Causal Effect of Economic Freedom on Female Employment & Education. *Contemporary Economic Policy,* 1–16. https://onlinelibrary.wiley.com/doi/abs/10.1111/coep.12674 (2024).

141. Debnam Guzman, J., Mabeu, M. C. & Pongou, R. Identity during a Crisis: COVID-19 and Ethnic Divisions in the United States. *AEA Papers and Proceedings* **112,** 319–324. https://www.aeaweb.org/articles?id=10.1257/pandp.20221115 (May 2022).

142. He, G. & Wang, S. Do College Graduates Serving as Village Officials Help Rural China? *American Economic Journal: Applied Economics* **9,** 186–215. https://www.aeaweb.org/articles?id=10.1257/app.20160079 (Oct. 2017).

143. Nguyen-Hoang, P. The Fiscal Effects of Tax Increment Financing on Rural School Districts: The Case of Iowa. *AERA Open* **7,** 2332858421991149. https://doi.org/10.1177/2332858421991149 (Jan. 1, 2021).

144. Howard, G. & Liebersohn, J. Why Is the Rent so Darn High? The Role of Growing Demand to Live in Housing-Supply-Inelastic Cities. *Journal of Urban Economics* **124,** 103369. https://www.sciencedirect.com/science/article/pii/S0094119021000516 (July 1, 2021).

145. Iyer, L., Mani, A., Mishra, P. & Topalova, P. The Power of Political Voice: Women's Political Representation and Crime in India. *American Economic Journal: Applied Economics* **4,** 165–193. https://www.aeaweb.org/articles?id=10.1257/app.4.4.165 (July 2012).

146. Jack, R., Halloran, C., Okun, J. & Oster, E. Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts. *American Economic Review: Insights* **5,** 173–190. https://www.aeaweb.org/articles?id=10.1257/aeri.20210748 (June 2023).

147. Jacob, B. A. School Board Elections Before, During, and After the COVID-19 Pandemic. *AERA Open* **10,** 23328584241291029. https://doi.org/10.1177/23328584241291029 (Nov. 1, 2024).

148. Jain, T. Common Tongue: The Impact of Language on Educational Outcomes. *The Journal of Economic History* **77,** 473–510. https://www.jstor.org/stable/26348367 (2017).

149. Jensenius, F. R. Development from Representation? A Study of Quotas for the Scheduled Castes in India. *American Economic Journal: Applied Economics* **7,** 196–220. https://www.aeaweb.org/articles?id=10.1257/app.20140201 (July 2015).

150. Charles, K. K., Li, Y. & Stephens Jr., M. Disability Benefit Take-Up and Local Labor Market Conditions. *The Review of Economics and Statistics* **100,** 416–423. https://doi.org/10.1162/rest_a_00692 (July 1, 2018).

151. Kumar, A. & Liang, C.-Y. Labor Market Effects of Credit Constraints: Evidence from a Natural Experiment. *American Economic Journal: Economic Policy* **16,** 1–26. https://www.aeaweb.org/articles?id=10.1257/pol.20200683 (Aug. 2024).

152. Luca, D. L., Owens, E. & Sharma, G. Can Alcohol Prohibition Reduce Violence against Women? *American Economic Review* **105,** 625–629. https://www.aeaweb.org/articles?id=10.1257/aer.p20151120 (May 2015).

153. Lundqvist, H., Dahlberg, M. & Mörk, E. Stimulating Local Public Employment: Do General Grants Work? *American Economic Journal: Economic Policy* **6,** 167–192. https://www.aeaweb.org/articles?id=10.1257/pol.6.1.167 (Feb. 2014).

154. Maestas, N., Mullen, K. J. & Strand, A. Disability Insurance and Health Insurance Reform: Evidence from Massachusetts. *American Economic Review* **104,** 329–335. https://pubs.aeaweb.org/doi/10.1257/aer.104.5.329 (May 1, 2014).

155. Martinez-Bravo, M. The Role of Local Officials in New Democracies: Evidence from Indonesia. *American Economic Review* **104,** 1244–1287. https://www.aeaweb.org/articles?id=10.1257/aer.104.4.1244 (Apr. 2014).

156. Martinez-Bravo, M. The Local Political Economy Effects of School Construction in Indonesia. *American Economic Journal: Applied Economics* **9,** 256–289. https://www.aeaweb.org/articles?id=10.1257/app.20150447 (Apr. 2017).

157. Mocan, H. N. & Bali, T. G. Asymmetric Crime Cycles. *The Review of Economics and Statistics* **92,** 899–911. https://doi.org/10.1162/REST_a_00048 (Nov. 1, 2010).

158. Nguyen, H.-L. Q. Are Credit Markets Still Local? Evidence from Bank Branch Closings. *American Economic Journal: Applied Economics* **11,** 1–32. https://www.aeaweb.org/articles?id=10.1257/app.20170543 (Jan. 2019).

159. Ortagus, J. C. *et al.* The Revenue Implications of Community Colleges' Reliance on Local Funding. *AERA Open* **10,** 23328584241258487. https://doi.org/10.1177/23328584241258487 (Nov. 1, 2024).

160. Pandey, P. Service Delivery and Corruption in Public Services: How Does History Matter? *American Economic Journal: Applied Economics* **2,** 190–204. https://www.aeaweb.org/articles?id=10.1257/app.2.3.190 (July 2010).

161. Perez, V., Benitez, J. & Seiber, E. Elasticity of Medicaid Access with Respect to State Financial Distress. *AEA Papers and Proceedings* **108,** 384–387. https://www.aeaweb.org/articles?id=10.1257/pandp.20181107 (May 2018).

162. Redding, S. J. & Sturm, D. M. The Costs of Remoteness: Evidence from German Division and Reunification. *American Economic Review* **98,** 1766–1797. https://www.aeaweb.org/articles?id=10.1257/aer.98.5.1766 (Dec. 2008).

163. Saiz, A. & Wachter, S. Immigration and the Neighborhood. *American Economic Journal: Economic Policy* **3,** 169–188. https://www.jstor.org/stable/41238097 (2011).

164. Shores, K. & Steinberg, M. P. Schooling During the Great Recession: Patterns of School Spending and Student Achievement Using Population Data. *AERA Open* **5,** 2332858419877431. https://doi.org/10.1177/2332858419877431 (July 1, 2019).

165. Slattery, C. & Zidar, O. Evaluating State and Local Business Incentives. *Journal of Economic Perspectives* **34,** 90–118. https://www.aeaweb.org/articles?id=10.1257/jep.34.2.90 (May 2020).

166. Baum-Snow, N. & Lutz, B. F. School Desegregation, School Choice, and Changes in Residential Location Patterns by Race. *American Economic Review* **101,** 3019–3046. https://www.aeaweb.org/articles?id=10.1257/aer.101.7.3019 (Dec. 2011).

167. Sosina, V. E. & Weathers, E. S. Pathways to Inequality: Between-District Segregation and Racial Disparities in School District Expenditures. *AERA Open* **5,** 2332858419872445. https://doi.org/10.1177/2332858419872445 (July 1, 2019).

168. Holzer, H. J., Hubbard, G. & Strain, M. R. Did Pandemic Unemployment Benefits Increase Unemployment? Evidence from Early State-Level Expirations. *Economic Inquiry* **62,** 24–38. https://onlinelibrary.wiley.com/doi/abs/10.1111/ecin.13180 (2024).

169. Sviatschi, M. M. US Criminal Deportations and Human Capital in Central America. *AEA Papers and Proceedings* **109,** 239–242. https://www.aeaweb.org/articles?id=10.1257/pandp.20191061 (May 2019).

170. Topalova, P. Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India. *American Economic Journal: Applied Economics* **2,** 1–41. https://www.aeaweb.org/articles?id=10.1257/app.2.4.1 (Oct. 2010).

171. Wang, F., Kis-Katos, K. & Zhou, M. Import Competition and the Gender Employment Gap in China. *Journal of Human Resources* **59,** 1830–1864. https://muse.jhu.edu/pub/19/article/940729 (2024).

172. Weber, M. & Baker, B. D. Does School Funding Matter in a Pandemic? COVID-19 Instructional Models and School Funding Adequacy. *AERA Open* **11,** 23328584251327581. https://doi.org/10.1177/23328584251327581 (Jan. 1, 2025).

173. Williams, J. Historical Lynchings and the Contemporary Voting Behavior of Blacks. *American Economic Journal: Applied Economics* **14,** 224–253. https://www.aeaweb.org/articles?id=10.1257/app.20190549 (July 2022).

174. Wilson, D. J. Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act. *American Economic Journal: Economic Policy* **4,** 251–282. https://www.aeaweb.org/articles?id=10.1257/pol.4.3.251 (Apr. 2012).

175. Gelman, A. & Stern, H. The Difference Between "Significant" and "Not Significant" Is Not Itself Statistically Significant. *The American Statistician* **60,** 328–331. https://doi.org/10.1198/000313006X152649 (Nov. 1, 2006).

176. Cumming, G. Replication and p Intervals: p Values Predict the Future Only Vaguely, but Confidence Intervals Do Much Better. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* **3,** 286–300 (July 2008).

177. Shoemate, M. *et al. OpenDP Library* version 0.12.1. 2025. https://github.com/opendp/opendp.

178. Dwork, C. & Roth, A. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science* **9,** 211–407. http://www.nowpublishers.com/articles/foundations-and-trends-in-theoretical-computer-science/TCS-042 (2013).

179. Kang, Y. *et al. Input Perturbation: A New Paradigm between Central and Local Differential Privacy* http://arxiv.org/abs/2002.08570. Pre-published.

180. Steinke, T. *Tight RDP & zCDP Bounds from Pure DP* DifferentialPrivacy.org. May 2024.

181. Stein, C. in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* 197–207 (University of California Press, Jan. 1, 1956). https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Third-Berkeley-Symposium-on-Mathematical-Statistics-and/chapter/Inadmissibility-of-the-Usual-Estimator-for-the-Mean-of-a/bsmsp/1200501656.

182. Berger, J. O. Admissible Minimax Estimation of a Multivariate Normal Mean with Arbitrary Quadratic Loss. *The Annals of Statistics* **4,** 223–226. https://www.jstor.org/stable/2958003 (1976).

Table 2: Effect of mechanism & other factors on strict epistemic parity over 10 replicates.

	(1)	(2)	(3)
(Intercept)	0.080	-0.007	0.597 ***
	(0.253)	(0.229)	(0.123)
Log epsilon	0.039 ***	0.039 ***	0.038 ***
	(0.003)	(0.003)	(0.003)
Gaussian mech. (zCDP)	0.034 ***	0.035 ***	0.034 ***
	(0.004)	(0.004)	(0.004)
Log epsilon x Gaussian mech. (zCDP)	-0.002 ***	-0.002 **	-0.002 **
	(0.001)	(0.001)	(0.001)
Num. vars. noised	0.014	0.013	-0.001
	(0.019)	(0.015)	(0.001)
Log sum (Sensitivity / mean)	-0.015	-0.017	-0.018 ***
	(0.011)	(0.009)	(0.004)
Original effect size		0.086	0.151
		(0.127)	(0.085)
Claim: insignificant		0.240 *	0.208 ***
		(0.093)	(0.058)
Claim: non-zero upper/lower bound		0.091	0.101 *
		(0.061)	(0.042)
Log regression sample size			-0.001
			(0.007)
Model degrees of freedom			-0.000
			(0.000)
Noised ind./treatment var.			-0.046
			(0.043)
Noised dep/outcome var.			-0.230 ***
			(0.046)
Region: City/municipality/village/zone/MSA/CBSA/AMC			0.055
			(0.040)
Region: NA			-0.029
			(0.052)
Region: Nation/country			-0.098
			(0.146)
Region: Organization (school, university, police unit)			0.041
			(0.083)
Region: State/district/province (1st division)			0.094 *
			(0.043)
Study FE	Yes	Yes	No
Implementation controls	Yes	Yes	Yes
Query type controls	Yes	Yes	Yes
N. obs.	34200	32946	32718
R squared	0.542	0.565	0.467
F statistic	391.081	417.832	1247.940

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are clustered by study.

Controls include number and type of component queries. Dummies for "Claim: sig. positive/negative" and "Region: County/tract/block/district/prefecture (below 1st division)" excluded. Study fixed effects excluded for Column 3, which includes covariates with little within-study variation. Includes 50 values of ϵ spaced evenly on a log scale in $[10^{-5}, 10^3]$ (both mechanisms).

A Methods

A.1 Replicated Findings

To find social science studies with working replication packages, we used 1) the Inter-University Consortium for Political and Social Research (ICPSR)'s replication package repository [44] and 2) Find Economic Articles with Data, a tool for searching English economics articles with accessible data/code supplements [45]. Both databases includes articles from journals with standardized data accessibility policies, including American Economic Association (AEA) journals, the *Review of Economics and Statistics*, and several other journals (Figure C.1). We searched for studies using aggregate statistics by filtering for articles whose title or abstract uses one of several common, mostly geographic, aggregation levels such as "county", "school", "neighborhood", or "organization". Appendix B.1 lists all our search terms.

We manually reviewed every article resulting from these searches, excluding studies that did not use aggregate statistics (e.g., studies exclusively analyzing microdata). We also excluded studies that did not use *personal data* (information about individuals or households). To construct a sample representative of contemporary research, we restricted the sample to articles published in the last 19 years that use personal data collected within the last 70 years (similar to the U.S. Census Bureau's policy of keeping data confidential for 72 years after collection).[9] We attempted to reproduce 166 studies that met these criteria[10], correcting the authors' code only as much as necessary to reproduce their results without error.[11] We were unable to replicate 73 studies, most often either because the replication package did not include the datasets necessary to reproduce the main results or because the datasets did not include enough information to implement differential privacy. Less often, our reproductions failed because of bugs or discrepancies in the replication packages.

Our current sample includes 93 successfully replicated studies [49, 54, 78, 85–174] , summarized in Table 1. For each article, we used the abstract and introduction to identify the key empirical claims made by the authors. We then identified the statistical estimate or estimates (hereafter *results*) supporting those claims—in our sample, all regression coefficients—and modified the authors' code to extract those key numerical results and related statistics (standard error, t-statistic, degrees of freedom, etc.).[12] For most studies, we reproduced 2–4 results (Table 1). Most results we reproduced (269 of 303) substantiate a key empirical claim mentioned in the article's abstract (Table 1).

[9] We also excluded 3 studies with code primarily written Matlab, which was not supported by our computing infrastructure. The final sample consisted of only Stata and R scripts, though a few studies included auxiliary Matlab scripts.

[10] The terms "repeatability", "reproducibility", and "replicability" are often interchanged [83]. Here, we use the term "reproduce" to mean achieving the same results with the same data and methods and "replicate" to mean achieving similar results using the same methods but different (noisy) data [84].

[11] We recorded the expected values of key results reported in the article and in every run, and we automatically checked those values against the results produced by our copy of the authors' code to check for errors introduced by our environment or modifications. All of our successful reproductions matched the authors' original results to same the level of numerical precision reported in their published article.

[12] To help standardize our analysis, we did not consider visual findings (multiple results represented by visual relation in a figure) [31]. Visual findings substantiated key claims in only a small handful of the studies we found.

A.2 Other Metrics

Sign parity: whether the counterfactual estimate $\hat{\beta^{\text{DP}}}$ has the same sign as the original estimate $\hat{\beta}$ at an equivalent level of significance for the test on $H_0 : \beta = 0$ vs. $H_A : \beta \neq 0$.

Sign reversal: whether the counterfactual estimate $\hat{\beta^{\text{DP}}}$ has the *opposite* sign as the original estimate $\hat{\beta}$ at an equivalent level of significance for the test on $H_0 : \beta = 0$ vs. $H_A : \beta \neq 0$.

Confidence interval coverage at α: whether the original estimate falls within the confidence interval of the counterfactual estimate at confidence level α.[13]

Bias in standardized effect size: the absolute difference between the counterfactual effect size and the original effect size. We follow Open Science Collaboration [4, Appendix A3] in converting each result to a common effect size statistic, the correlation coefficient r computed from the t-statistic and residual degrees of freedom ν for each result:

$$r = \sqrt{\frac{t^2/\nu}{t^2/\nu + 1} + 1} \left(\text{sign}(\hat{\beta}) \text{sign}(\hat{\beta^{\text{DP}}}) \right) . \tag{1}$$

r is coded as negative if the sign of the counterfactual estimate does not match the sign of the original estimate.

A.3 Differential Privacy

To investigate the counterfactual where the aggregate statistics in these studies were released with differential privacy, we injected noise into all statistics about personal data used to produce the key results in our sample.

We compare two approaches: 1) pure differential privacy with the Laplace mechanism [20], a simple randomized mechanism for releasing statistics; and 2) zero-concentrated differential privacy (zCDP) [46, 47] with the Gaussian mechanism, a relaxation of differential privacy with improved utility.[14] These are two simple, popular, and generalizable DP mechanisms data curators might implement, though there are more optimal techniques for utility, such as data-adaptive mechanisms [55].

In pure differential privacy, the parameter ϵ measures how much the inclusion of information about an individual in the dataset increases the risk that their personal information is revealed by published statistics (Def. 1).

Each study in our sample uses a set of ν variables (or covariates) $F(D)_{k\times\nu} = (f_1(D), f_2(D), \ldots f_\nu(D))$: statistical queries $f_j : \mathcal{D}^n \to \mathbb{R}^k$ over a confidential dataset of personal information D about n individuals.[15] Each statistical query $f_j(D) = (f_j(d_1), f_j(d_2), \ldots, f_j(d_k))$ is provided for k different disjoint subsets $d_i \in D$ (e.g., the total population of each region in each year).

[13]Authors often rely on significance thresholds to make claims, but statistical significance is often an imprecise and arbitrary measure for epistemic justification. Often, only a small change is required to move an estimate from one significance level to another [175]. Confidence intervals may give much better information about replicability than p-values, which can vary widely over replications of the same statistical test [176].

[14]Specifically, we used the implementation of the Laplace and Gaussian mechanisms in the `opendp` Python library [177].

[15]The original microdata are typically not provided in the replication package to preserve confidentiality; we only have access to the statistics $F(D)$ used by the original authors.

Definition 1 [20]. Let $M : \mathcal{D}^n \to \mathcal{Y}$ be a randomized algorithm. M is *ϵ-(pure) differentially private* if for all pairs of datasets $D, D' \in D^n$ which differ in only one entry and all possible outputs $T \subseteq \mathcal{Y}$:

$$\text{Pr}[M(D) \in T] \leq e^{\epsilon}\text{Pr}[M(D') \in T].$$

Pure DP with the Laplace mechanism. To simulate releasing $F(D)_{k\times\nu}$ with ϵ-(pure) differential privacy, we apply the Laplace mechanism to each query $f_j(D)$ (Def. 3). We rely on basic composition [178, Corollary 3.15] to release all the queries $F(D)_{k\times\nu}$ with ϵ-DP: we use $\frac{\epsilon}{\nu}$-DP mechanisms $M_j^{Lap}(D)$ for each query f_j such that $M^{\text{Lap}}(D) = \left(M_1^{Lap}(D), M_2^{Lap}(D), \ldots, M_\nu^{Lap}(D)\right)$ is ϵ-(pure) DP.

This is an "input perturbation" approach to regression modeling [179]. Input perturbation describes the common scenario where researchers run regressions on public statistics released with DP (e.g., from the Decennial Census or Facebook URLs dataset [7, 76]). Less noisy estimates could potentially be obtained through "output" or "objective" perturbation, in which a trusted data curator centrally mediates the analysis process by instead adding noise to the outputs (coefficients) of regression models before returning them to analysts [179], but this setup is less common and requires more effort from both data curators and researchers [38].

For some but not all studies, it is possible that the same individual could contribute to multiple subsets (usually statistics from different time periods). For simplicity, we only protect privacy with respect to an individual's contribution to one subset d_i; in general, then, our mechanisms are thus ϵ differentially private with respect to the contribution of each individual within every subset. Since individuals typically contribute to only one geographic region, our "unit" of privacy is usually (individual)-(time period) [23]. With that assumption, $\Delta f_j(D) = \Delta f_j(d_i)\ \forall d_i \in D$ (equivalent to parallel composition).
Definition 2 [178]. The *l_p-sensitivity* of a function $f : \mathcal{X}^n \to \mathbb{R}^k$ is:

$$\Delta f = \max_{D,D'} ||f(D) - f(D')||_p,$$

where D, D' are datasets which differ in exactly one entry.
Definition 3 [178]. Let $f : \mathcal{D}^n \to \mathbb{R}^k$. The *Laplace mechanism* is defined as

$$M^{Lap}(D) = f(D) + (Y_1, \ldots, Y_k)\,,$$

where the Y_i are independent random variables drawn from the Laplace distribution $Y_i \sim Lap\left(\frac{\Delta f}{\epsilon}\right)$ (Def 6), and Δf is the l_1-sensitivity of the query function f (Def. 2).

zCDP with the Gaussian mechanism. Zero-concentrated differential privacy (zCDP; Def. 4) is a relaxation of pure differential privacy that provides better accuracy [46, 47]. The Gaussian mechanism (Def. 5) satisfies zCDP. This mechanism was used to release statistics from the 2020 Decennial Census [7].

zCDP uses a different privacy parameter, ρ. A mechanism that satisfies pure DP also satisfies cZDP; specifically, if M satisfies ϵ-DP, then M satisfies $\left(\epsilon\frac{\exp(\epsilon)-1}{\exp(\epsilon)+1} < \frac{\epsilon^2}{2}\right)$-zCDP [47, 180]. We use this bound for comparison between the two mechanisms hereafter. zCDP has the same basic composition property as pure DP [47, Lemma 1.7], so we again apply $\frac{\rho}{\nu}$-zCDP mechanisms $M_j^{\text{Gauss}}(D)$ such that $M^{\text{Gauss}}(D) = \left(M_1^{Gauss}(D), M_2^{Gauss}(D), \ldots, M_\nu^{Gauss}(D)\right)$ is ρ-zCDP.

Definition 4 [47]. Let $M : \mathcal{D}^n \to \mathcal{Y}$ be a randomized algorithm. M is ρ-zCDP if for all pairs of datasets $D, D' \in D^n$ which differ in only one entry and all $\alpha \in (1, \infty)$:

$$D_\alpha(M(D)||M(D')) \leq \rho\alpha,$$

where $D_\alpha(M(D)||M(D'))$ is the α-Rényi divergence between the distributions of $M(D)$ and $M(D')$ (Def. 7).

Definition 5 [47]. Let $f : \mathcal{D}^n \to \mathbb{R}^k$. The *Gaussian mechanism* is defined as

$$M^{Gauss}(D) = f(D) + (Z_1, \ldots, Z_k)\,,$$

where the Z_i are independent random variables drawn from the Gaussian distribution $Z_i \sim \mathcal{N}\left(\frac{(\Delta f)^2}{2\rho}\right)$, and Δf is the l_2-sensitivity of the query function f (Def. 2).

Sensitivity and post-processing. For each study in our sample, we identified all the statistics about personal data used to produce the key results. For each statistic, we defined the global sensitivity of the query used to produce each individual statistic (Def. 2). When the global sensitivity of a query was undefined (for a ratio of counts, for example), we first deconstructed the query into component queries with defined sensitivity (e.g., the numerator and denominator counts), applied the DP mechanism to each component independently, then reconstructed the query.[16] Some variables are used as regression weights and must not be negative; for those variables, we clipped the noisy values to 0. Differential privacy is preserved through these post-processing steps [178, Proposition 2.1].

Because almost all of the datasets used in our sample contain only aggregate statistics, and not the confidential microdata used to construct them, we make several simplifying assumptions about the data curator that likely differ from a real-world deployment of DP.

1. For a few statistics used in our sample, the authors' replication package did not provide enough information to define sensitivity directly or reconstruct the statistical query. For some queries on continuous data (mostly about income or wages), we were able to make a simplifying assumption about the sensitivity (e.g., that income does not exceed three standard deviations above the mean). We did not apply any privacy protection to the remaining statistics (16.3% of statistics in our sample).

2. We assume that individuals have equal influence over statistics. In reality, particularly for survey-based estimates, some subgroups may be weighted more heavily than others. A real-world deployment of differential privacy must take this into account [see, e.g., 61, 62].

3. Statistics are likely to be released as part of larger datasets, and data curators may have to split privacy loss budgets amongst many other queries. We only distribute the privacy budget ϵ or ρ over the statistics used within each study, probably resulting in less noise than would

[16] We included these pre- and post-processing computations even in control runs where no noise was injected, checking to ensure that our reconstructions matched the original data.

be injected in practice.

4. In our mechanisms, each statistic receives an equal share of the privacy budget. In real-world deployments such as the U.S. Decennial Census, data curators may tune budgets more carefully to preference certain critical statistics [14], adding or reducing noise depending on policy priorities.

5. The statistics used in these studies may be derivatives or linkages of other public datasets produced by other data curators; because these precursor datasets and pre-processing steps are usually not accessible, we apply differential privacy to the combined set of statistics used directly by the authors, as if all the data curators shared a single total privacy budget ϵ or ρ for each study.

6. We assume that the statistics used in our sample are not already subject to formal privacy protections. In reality, the data that produced these statistics may have been subject to cell suppression, random swapping, censoring, or other disclosure avoidance techniques before publication.

A.4 Additive Data Error

To compare the effects of DP mechanisms to the possible impacts of existing data error, we simulate possible alternative versions of the dataset by drawing counterfactual observations from a normal distribution conditioned on the original statistics.

Variance. Unfortunately, the replication datasets rarely include data-based variance estimates. Thus, for each observed statistic $x_{ij} = f_j(d_i)$, we assume some coefficient of variation c_{ij} and simulate sampling variances $v_{ij} = (c_{ij}x_{ij}^{1-b})^2$ in proportion to the observed statistic. The parameter b models the rate at which the standard deviation $c_{ij}x_{ij}$ diminishes with x_{ij}; by default, $b = 0$ (constant returns). We test a range of coefficients of variation $c \in \mathbb{R}^+$ and $b \in \mathbb{R}$.

Shrinkage. Let μ represent the true statistics. Simply drawing replicates from

$$\mu | x \sim \mathcal{N}(x, \text{diag}(v)) \tag{2}$$

is inadmissable for the true population statistics μ [181]. Cui *et al.* [65] suggest two admissable constructions using shrinkage estimation from multi-level empirical Bayesian modeling. Following Cui *et al.* [65], we simulate replicates with the general form

$$\mu_{ij} | x \sim \mathcal{N}\left((1 - B_{ij})x_{ij} + B_{ij}\beta_j, (1 - B_{ij})v_{ij}\right), \tag{3}$$

where $\beta_j, B_{ij} \in [0, 1]$ are functions of x_{ij} and v_{ij} [65, 67]. This method adjusts the counterfactual estimate to account for a baseline β_j with variance reduced by $100B_{ij}\%$. The Hudson-Berger construction [66, 182] uses $\beta = 0$ and

$$B_{ij}^{HB} = \min\left(1, \frac{(k-2)/v_{ij}}{\sum_{m=1}^{k}(x_{mj}/v_{mj})^2}\right).$$

The Morris-Lysy construction [67] uses $\beta_j = \bar{x}_j$ and

$$B_{ij}^{ML} = \frac{v_{ij}}{v_{ij} + \bar{v}_j^H (1 - \hat{B}_j^H)/\hat{B}_j^H},$$

where $\bar{v}_j^H = k / \sum_{i=1}^k v_{ij}^{-1}$ is the harmonic mean of v_j, $\hat{B}_j^H = (k-3)/(k-1)\hat{\sigma}^2$, and $\hat{\sigma}^2 = (k-1)^{-1} \sum_{i=1}^k (x_{ij} - \bar{x}_j)^2 / v_{ij}$ is the mean squared error in the observed statistic x_j.

Generally, the Hudson-Berger construction imposes more shrinkage for large v_{ij}, while Morris-Lysy imposes more shrinkage for smaller v_{ij}. We also present results for the inadmissable no-shrinkage construction (Eq. 2) for comparison.

B Supplementary Methods

B.1 Search Terms

For each data source, we reviewed all results with abstract or title containing one of the following keywords: “state”, “county”, “city”, “municipality”, “district”, “province”, “block”, “commuting zone”, “statistical area”, “prefecture”, “tract”, “college”, “university”, “school”, “neighborhood”.

B.2 Definitions

Definition 6 [178]. The *Laplace distribution* (centered at 0) with scale b is the distribution with probability density function:

$$Lap(x|b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right).$$

Definition 7 [47]. Let P and Q be probability distributions on Ω. For $\alpha \in (1, \infty)$, the *Rényi divergence* of order α between P and Q is:

$$D_\alpha = \frac{1}{\alpha - 1} \log \mathbb{E}_{x \sim P} \left[\left(\frac{P(x)}{Q(x)} \right)^{\alpha - 1} \right]$$

B.3 Other Metrics

Significance match. Following Williams *et al.* [53], we define significance match as the rate at which the significance level ($p < 0.01$, $p < 0.05$, $p < 0.1$, or insignificant) of the noisy estimate matches that of the original estimate, regardless of sign or magnitude.

Sign match. Following Williams *et al.* [53], we define sign match as the rate at which the counterfactual estimate has the same sign as the original estimate, regardless of statistical significance.

Effective sample size. Following Williams *et al.* [39], we define the effective sample size of the counterfactual estimate as the number of observations n_{ESS} such that the counterfactual variance $\text{Var}(\hat{\beta^{\text{DP}}})$ would equal the original variance $\text{Var}(\hat{\beta})$:

$$n_{\text{ESS}} = n \frac{\text{Var}(\hat{\beta})}{\text{Var}(\hat{\beta^{\text{DP}}})}. \tag{4}$$

C Supplementary Figures & Tables

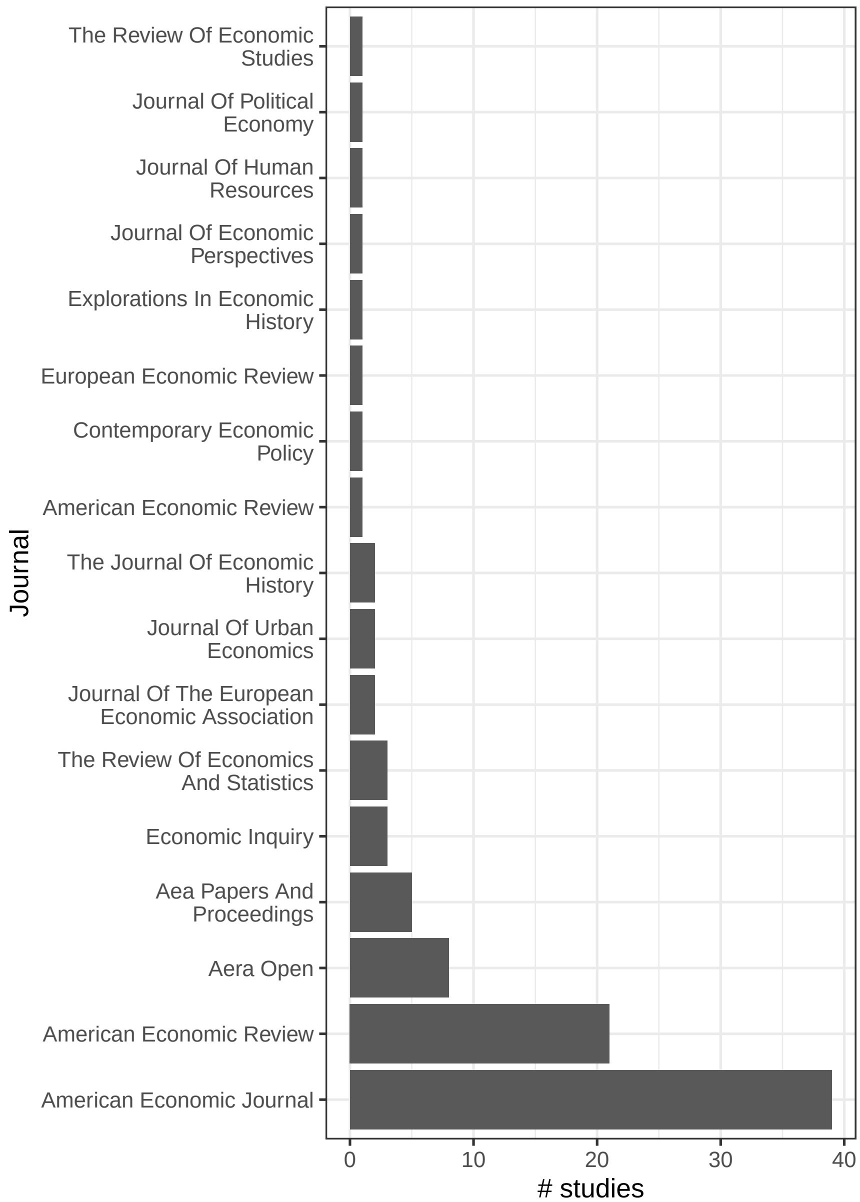


Figure C.1: Number of studies per journal.

Table C.1: Effect of mechanism and other factors on log RMSD of noise added to each component statistical query.

	(1)	(2)	(3)	(4)
(Intercept)	2.725 ***	4.857 ***	3.246 ***	1.958 ***
	(0.567)	(0.122)	(0.153)	(0.036)
Log epsilon	-0.975 ***		-0.975 ***	-0.974 ***
	(0.007)		(0.007)	(0.007)
Gaussian mech. (zCDP)	-0.989 ***		-0.968 ***	-0.968 ***
	(0.102)		(0.082)	(0.082)
Log epsilon x Gaussian mech. (zCDP)	0.129 ***		0.129 ***	0.129 ***
	(0.001)		(0.001)	(0.001)
Log sensitivity		0.812 ***	0.846 ***	0.841 ***
		(0.061)	(0.042)	(0.082)
Log sensitivity x Gaussian mech. (zCDP)		0.102	0.034	0.034
		(0.084)	(0.029)	(0.029)
Study FE	No	No	No	Yes
N. obs.	65991	65991	65991	65991
R squared	0.716	0.242	0.957	0.977
F statistic	55340.427	10541.884	293704.836	30519.599

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are clustered by study.

Table C.2: Effect of mechanism & other factors on strict epistemic parity over 10 replicates.

	(1)	(2)	(3)
(Intercept)	0.595 ***	-0.340	0.293 **
	(0.056)	(0.230)	(0.113)
Log coeff. of variation (a)	-0.058 ***	-0.058 ***	-0.058 ***
	(0.004)	(0.005)	(0.004)
Morris-Lysy construction	-0.295 ***	-0.285 ***	-0.283 ***
	(0.026)	(0.026)	(0.025)
Num. vars. noised	0.009	0.055 *	-0.001
	(0.011)	(0.023)	(0.001)
Log regression sample size		0.003	0.012
		(0.006)	(0.007)
Model degrees of freedom		0.000	-0.000
		(0.000)	(0.000)
Noised ind./treatment var.		-0.150 *	-0.059
		(0.073)	(0.033)
Noised dep/outcome var.		-0.067	-0.180 ***
		(0.036)	(0.034)
Region: City/municipality/village/zone/MSA/CBSA/AMC		0.161	0.078
		(0.123)	(0.040)
Region: State/district/province (1st division)		0.132	0.031
		(0.088)	(0.039)
Region: NA			-0.023
			(0.052)
Region: Nation/country			-0.120
			(0.197)
Region: Organization (school, university, police unit)			0.170 **
			(0.056)
Original effect size			0.193 **
			(0.070)
Claim: insignificant			0.204 ***
			(0.043)
Claim: non-zero upper/lower bound			0.069
			(0.040)
Study FE	Yes	Yes	No
Implementation controls	Yes	Yes	Yes
Query type controls	No	Yes	Yes
N. obs.	10605	10150	10150
R squared	0.394	0.412	0.333
F statistic	71.040	68.954	240.332

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are clustered by study.

Controls include number and type of component queries. Dummies for Laplace mechanism (pure DP), “Cmd: reg, xtreg”, “Claim: sig. positive/negative”, and “Region: County/tract/block/district/prefecture (below 1st division)” excluded. Study fixed effects excluded for Column 3, which includes covariates with little within-study variation. Includes all experiments for additive data error (both constructions), $c \in \{10^{-3}, 10^{-2}, 0.05, 0.1, 0.2, 0.5\}$ with $b = 0$.

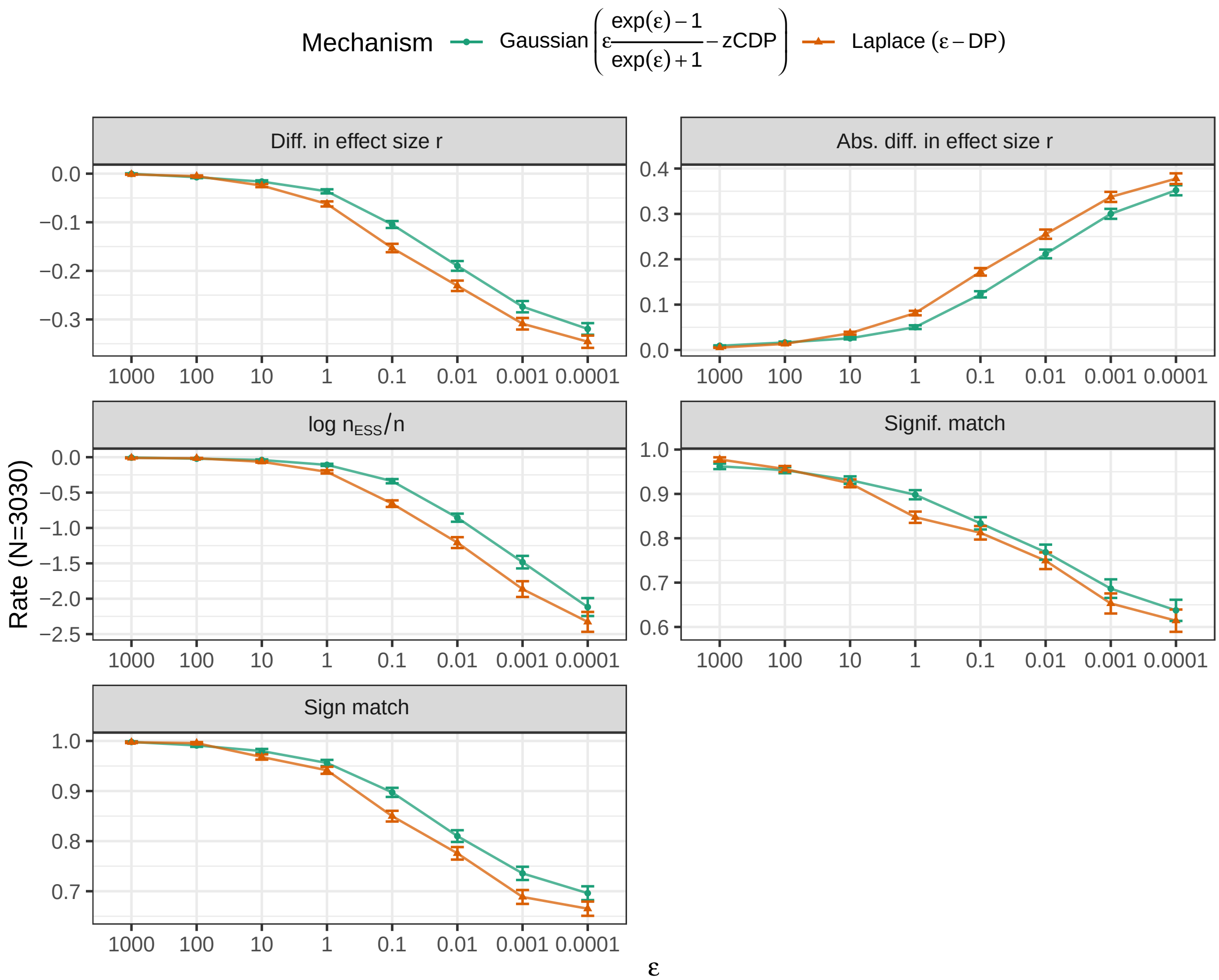


Figure C.2: Impacts of various settings of the privacy parameter ϵ over 10 runs of DP, averaged over all results. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$.

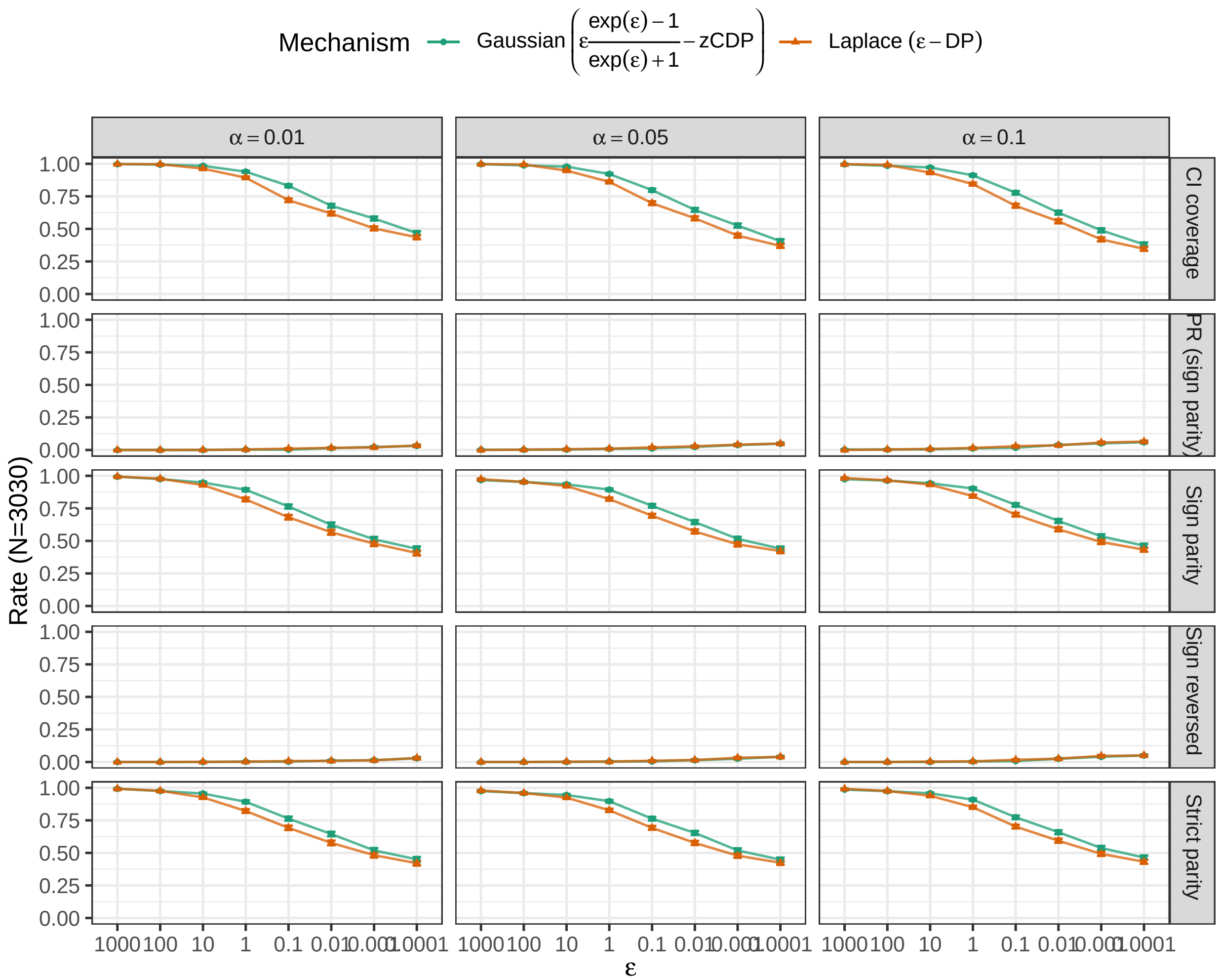


Figure C.3: Impacts of various settings of the privacy parameter ϵ at various significance levels α over 10 runs of DP, averaged over all results. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$.

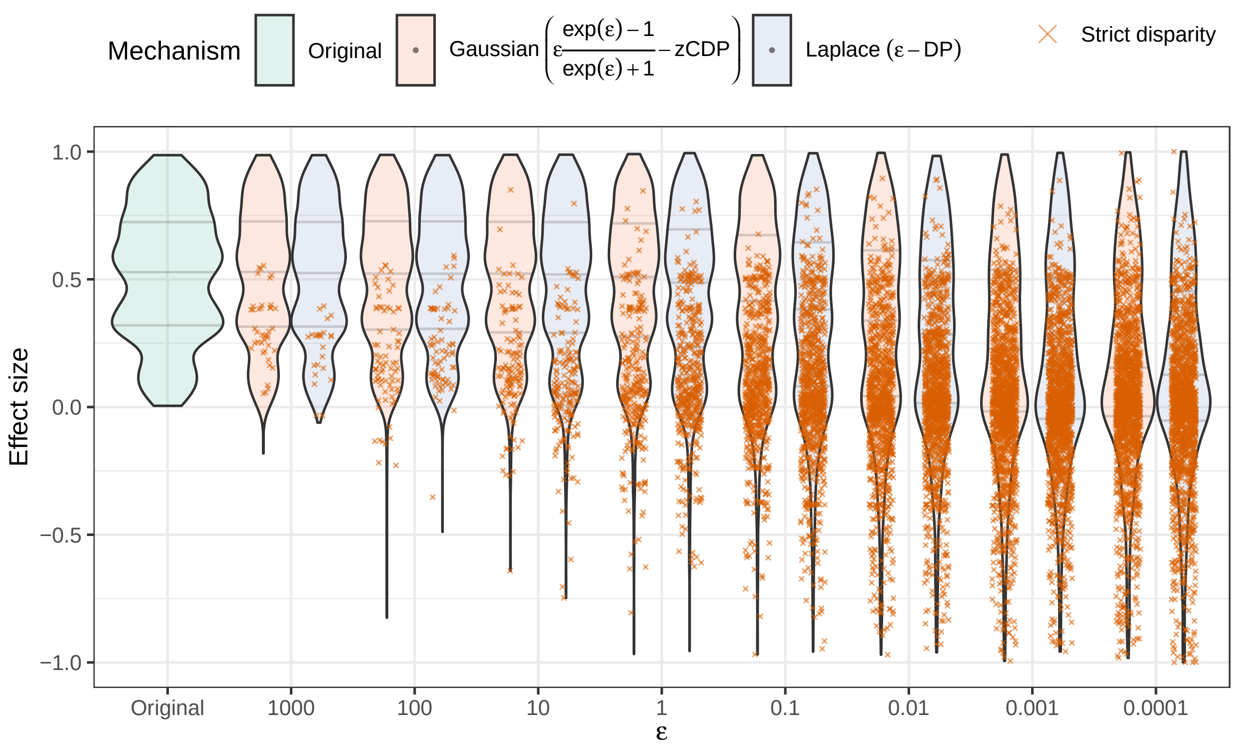


Figure C.4: Distribution of standardized effect sizes by study at different settings of the privacy parameter ϵ, each repeated 10 times. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$.

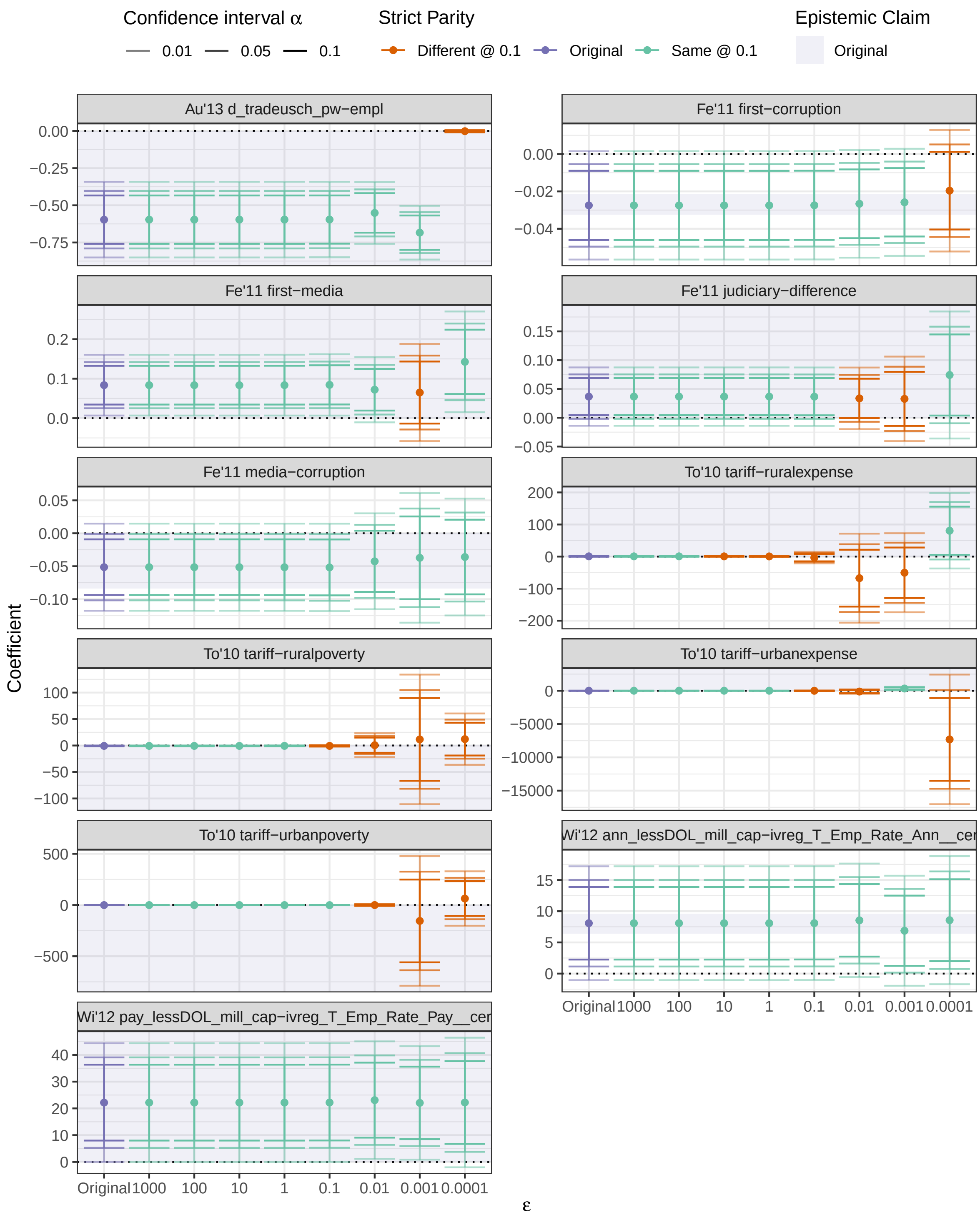


Figure C.5: Estimates from the first run of the ϵ-DP Laplace mechanism for a sample of studies [78, 131, 170, 174]. Epistemic parity colored based on confidence level $\alpha = 90$. The shaded region indicates the parity region defined by the original epistemic claim (e.g., that the coefficient is greater than zero).

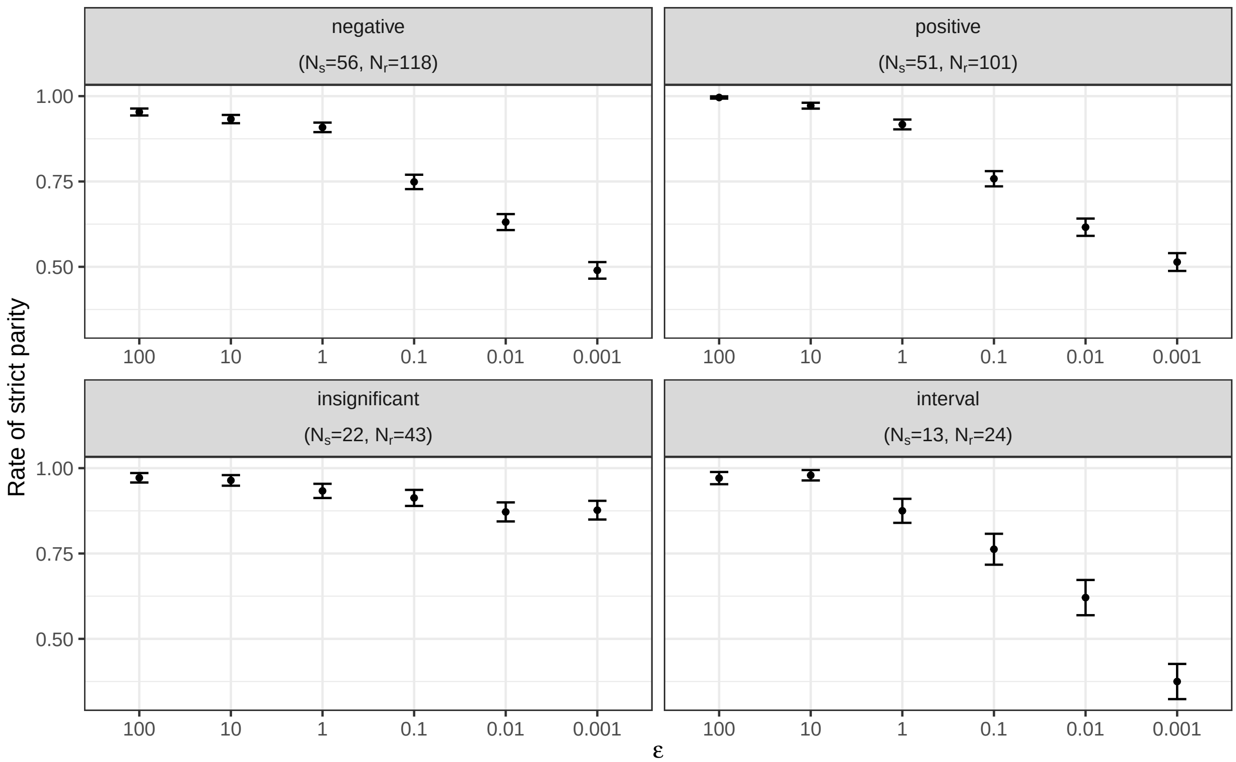


Figure C.6: Average epistemic parity vs. the privacy parameter ϵ across different types of epistemic claims over 10 runs of the ϵ-DP Laplace mechanism. Epistemic parity averaged across noised statistics. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$. N_s is the number of studies in each category; N_r is the number of results.

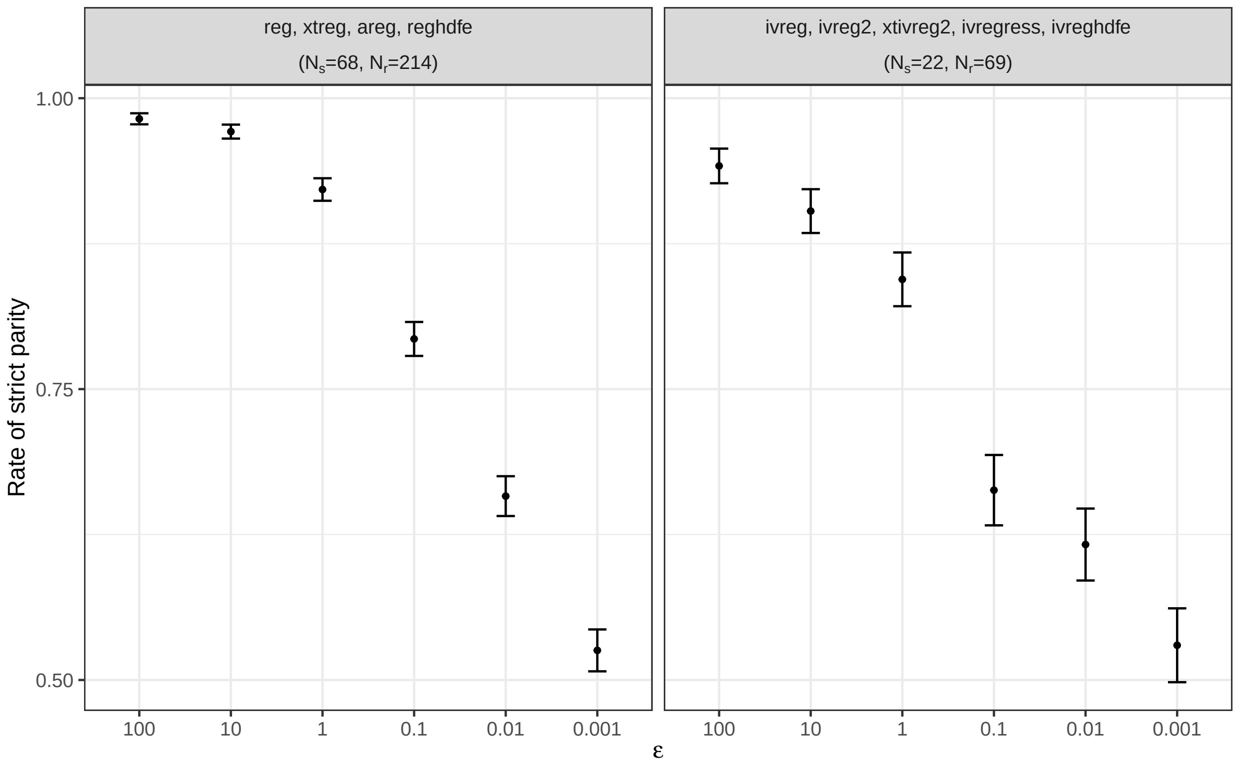


Figure C.7: Average epistemic parity for different Stata regression commands and different values of the privacy parameter ϵ over 10 runs of the ϵ-DP Laplace mechanism. Epistemic parity averaged across noised statistics. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$. N_s is the number of studies in each category; N_r is the number of results.

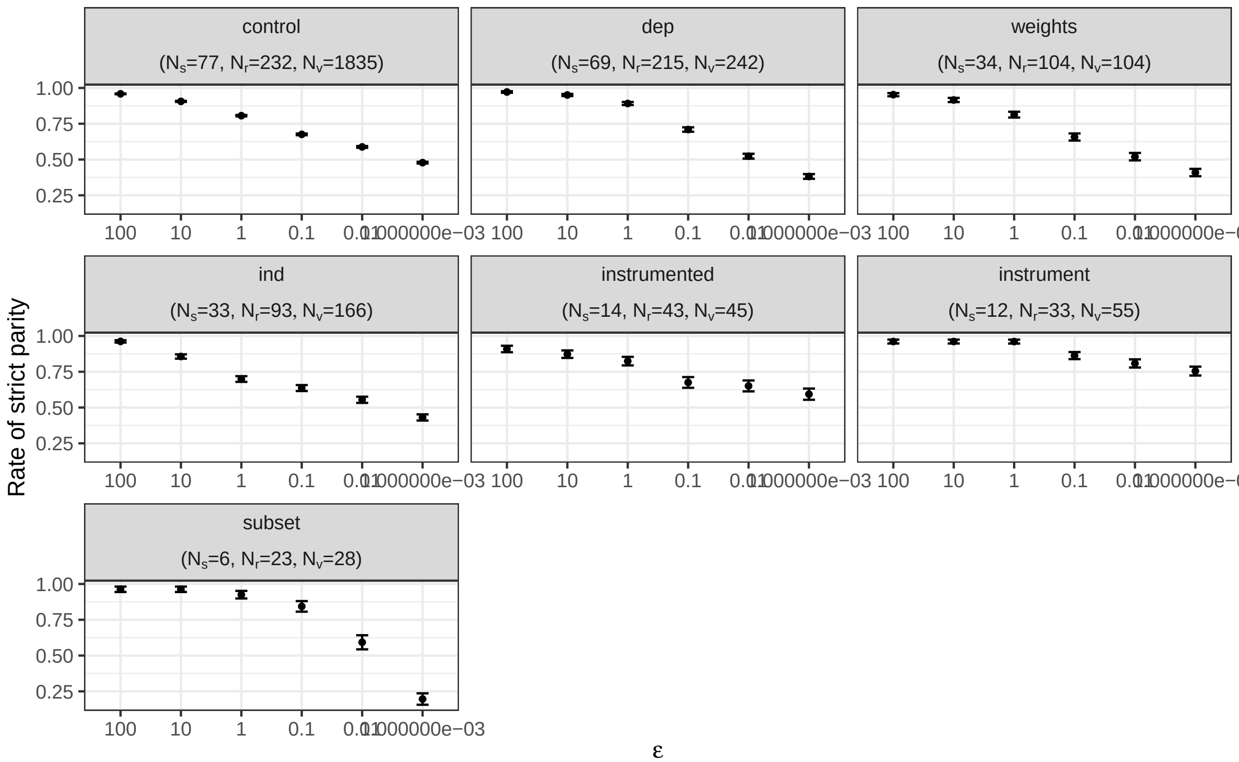


Figure C.8: Average epistemic parity vs. the privacy parameter ϵ over 10 runs of the ϵ-DP Laplace mechanism. Epistemic parity averaged across noised statistics and grouped by the role of each statistic in the regression analysis. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$. N_s is the number of studies in each category; N_r is the number of results; N_v is the number of variables.

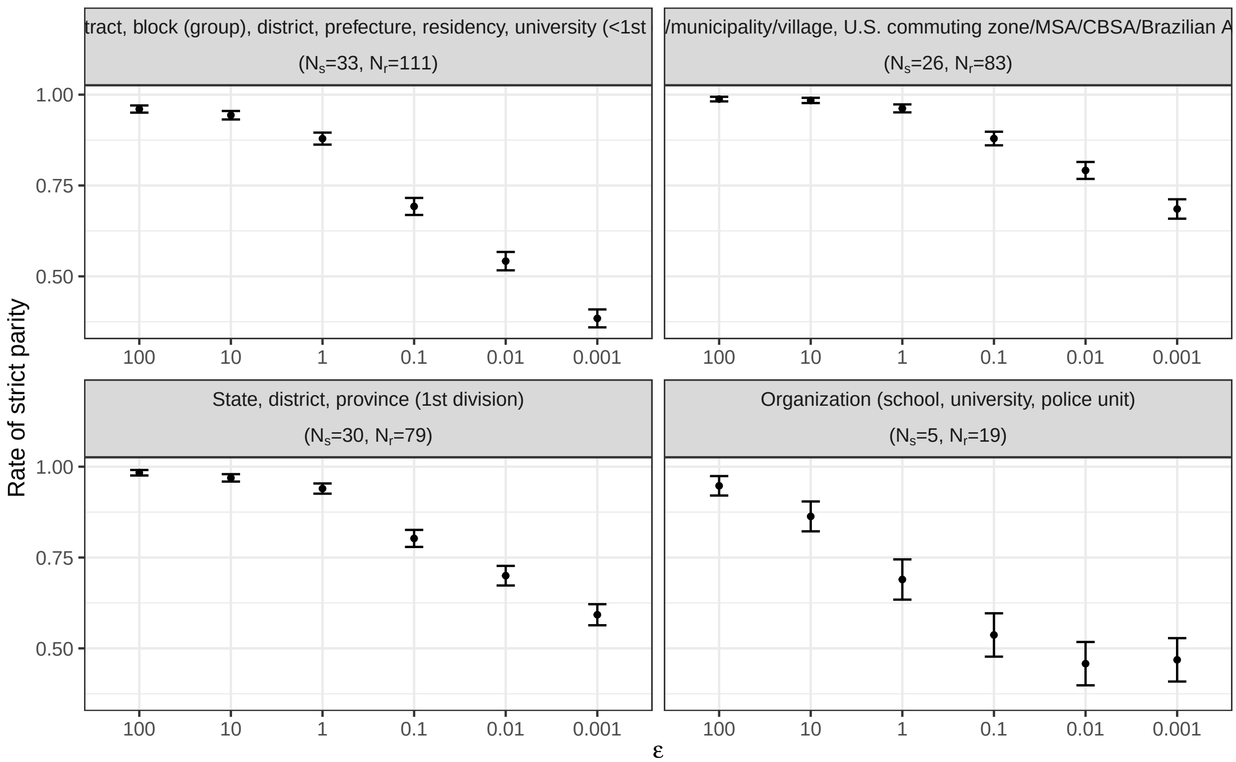


Figure C.9: Average epistemic parity for different aggregation levels and different values of the privacy parameter ϵ over 10 runs of the ϵ-DP Laplace mechanism. Epistemic parity averaged across noised statistics. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$.

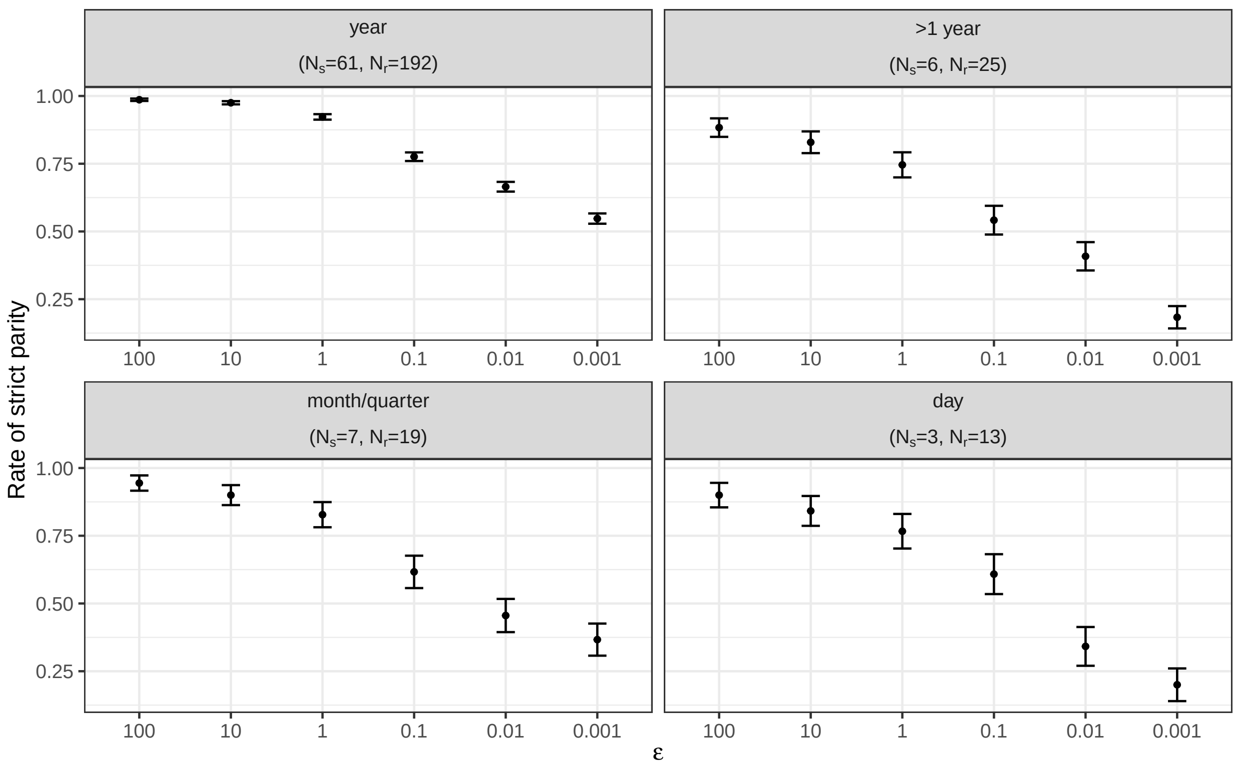


Figure C.10: Average epistemic parity for different aggregation levels and different values of the privacy parameter ϵ over 10 runs of the ϵ-DP Laplace mechanism. Epistemic parity averaged across noised statistics. NA indicates aggregations with no time index.Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$.

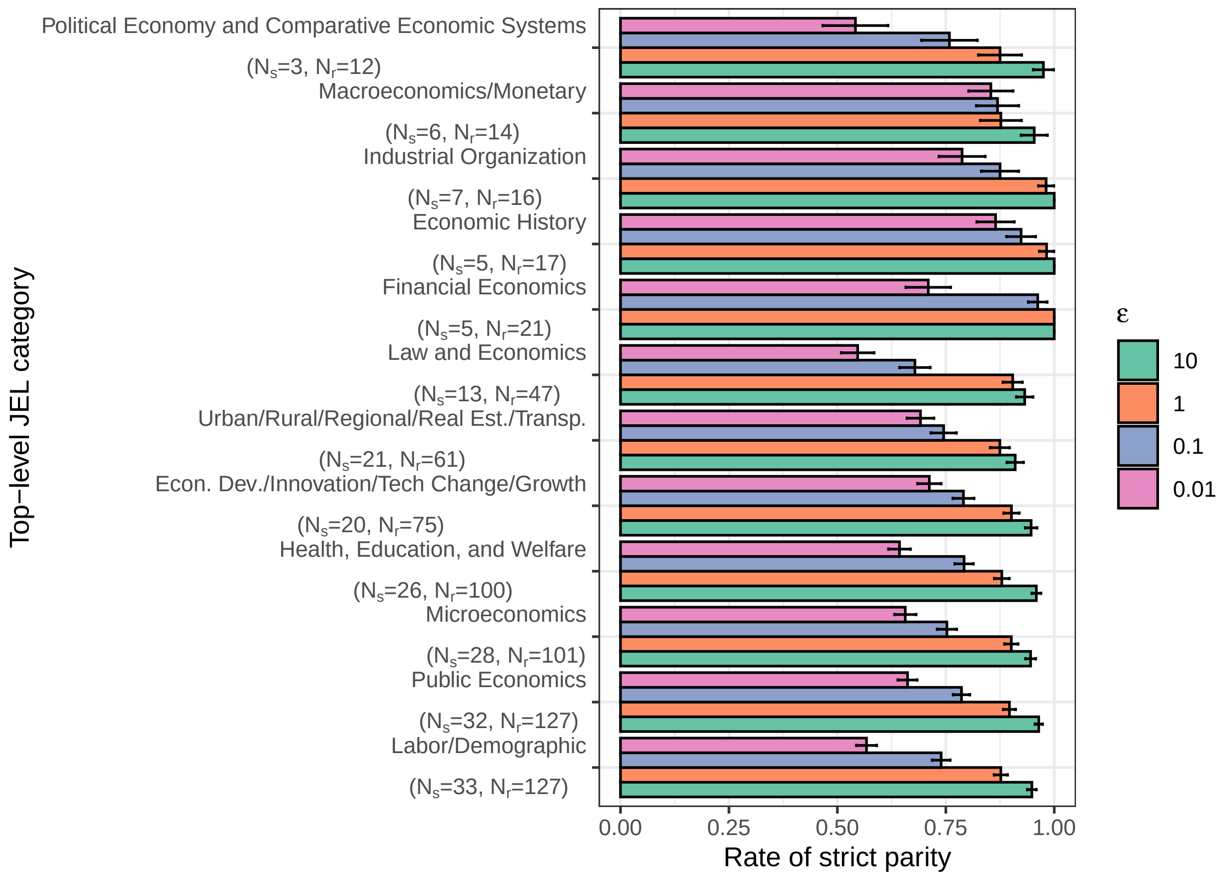


Figure C.11: Average epistemic parity for different JEL categories and different values of the privacy parameter ϵ over 10 runs of the ϵ-DP Laplace mechanism. Studies may appear in multiple JEL categories. Epistemic parity averaged across results. Excluding categories with less than 10 results. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$.

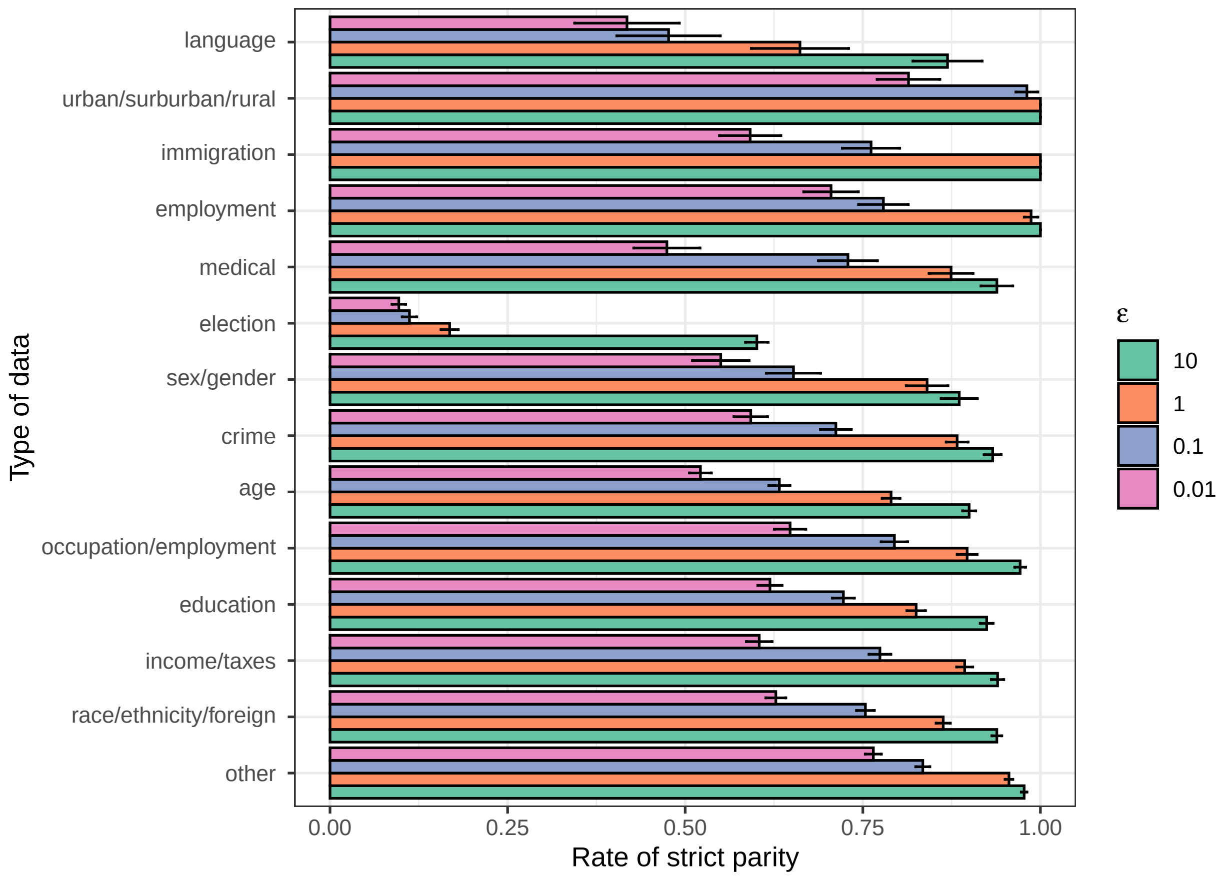


Figure C.12: Average epistemic parity for different types of statistics and different values of the privacy parameter ϵ over 10 runs of the ϵ-DP Laplace mechanism. Epistemic parity averaged across noised statistics. Excluding categories with less than 10 results. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$.

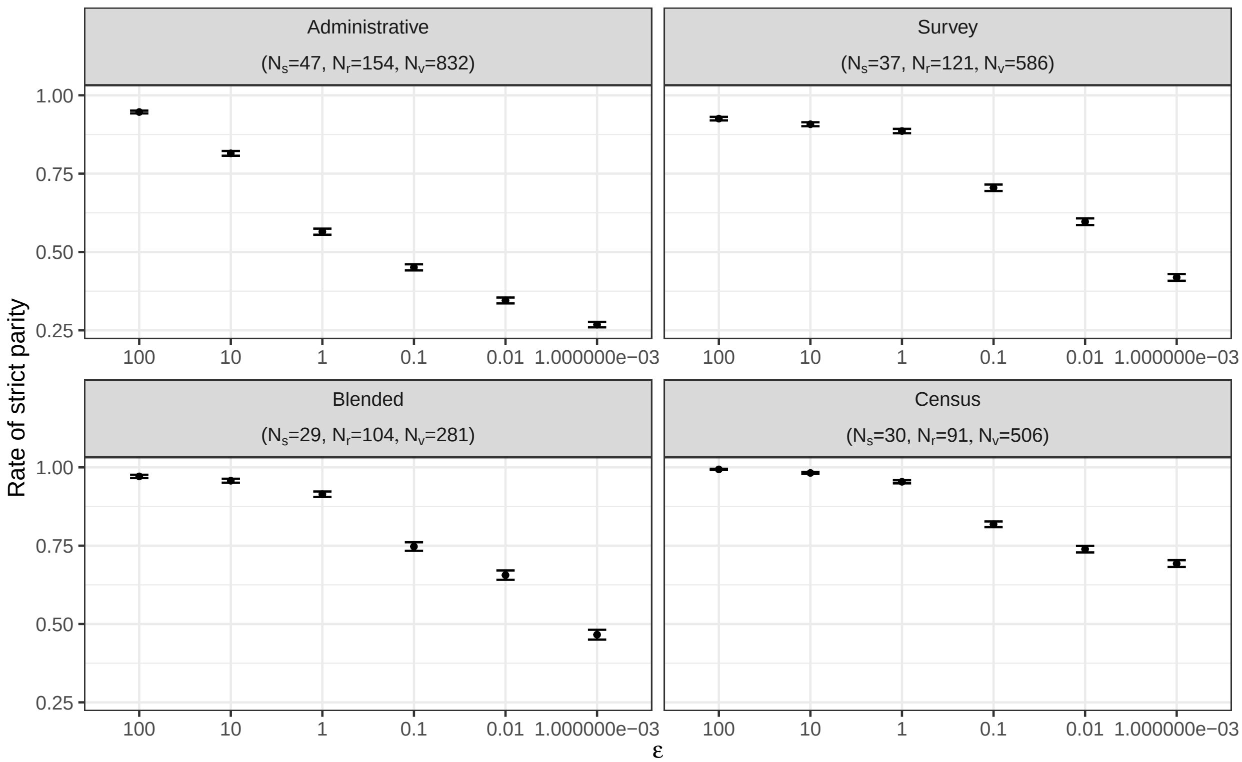


Figure C.13: Average epistemic parity for different types of data sources and different values of the privacy parameter ϵ over 10 runs of the ϵ-DP Laplace mechanism. Epistemic parity averaged across noised statistics. NA indicates unknown or unclear data products. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$. N_s is the number of studies in each category; N_r is the number of results; N_v is the number of variables.