



Reproducibility of social science research using aggregate statistics with noise infused for differential privacy

Ryan Steed^a , Eduardo Abraham Schnadower Mustri^a, and Alessandro Acquisti^{b,1}

Affiliations are included on p. 11.

Edited by Jerome P. Reiter, Duke University, Durham, NC; received February 14, 2026; accepted August 24, 2026, by Editorial Board Member Kenneth W. Wachter

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 the quality and feasibility of social science research. We select a benchmark of empirical findings from 93 published social science studies involving regression analyses over aggregate statistics. We evaluate whether their findings replicate on privacy noise-infused data. Under privacy budgets typical in industry, around 91% of simulated findings still support the original claims at significance 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 nonsampling 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.

differential privacy | reproducibility | public statistics

Industry and government organizations from Meta to the US Census Bureau have turned to new methods to strengthen privacy protections in data releases following concerns about attacks aided by commercial data and increased computing power (1–7).*

Applications of these modern privacy protections, however, have met with controversy (16). One of the most popular approaches, differential privacy (DP) (17, 18), mathematically guarantees that any one person's data has a limited effect on statistical outputs, usually by infusing statistical noise when sharing statistics about sensitive personal data. After recent policies—including presidential executive orders—incentivized the adoption of DP and similar methods (19, 20), 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” (21), “greatly reduced social knowledge” (22), and distorted government decision-making (23, 24). On the other hand, some scholars, public officials, and privacy advocates argue that failing to deploy noise infusion or other modern protections will forfeit opportunities for research on previously inaccessible private data, and may potentially deteriorate public trust and future data quality (6, 25, 26). 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 regression analysis findings from 93 empirical studies published in economics and social science journals, and then compared their published findings to a counterfactual version with simulated differentially private aggregate statistics. We evaluated epistemic parity, the principle that published findings should be replicable with a noise-infused dataset (27). Assuming researchers do not alter their methods to account for added noise, would social science studies reach the same conclusions if they relied on aggregate statistics infused with noise for privacy?

We find that at the scale of privacy budgets typical in industry ($\epsilon = 1$; a measure of how much an individual's inclusion increases their risk of disclosure), around 91% of

Significance

To help preserve privacy, the US Census Bureau, Meta, and other prominent organizations have added statistical randomness, or noise, to statistics about people before public release. This technique has triggered heated debate among social scientists about the potentially deleterious impacts of random noise on scientific findings. Replicating 93 social science studies involving regression analyses of aggregate statistics, we find that, under levels of noise commonly used in industry, the impacts of differential privacy on reproducibility are limited. They are smaller than impacts due to typical levels of data error, and smaller than rates of disparity typically observed in replication studies. Our study estimates key trade-offs that could inform policymakers and data curators designing the next generation of privacy-preserving technologies.

Author contributions: R.S. and A.A. designed research; R.S. and E.A.S.M. performed research; R.S. and E.A.S.M. analyzed data; and R.S. and A.A. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. J.P.R. is a guest editor invited by the Editorial Board.

Copyright © 2026 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: acquisti@mit.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2605550123/-DCSupplemental>.

Published October 9, 2026.

*There are many empirical examples of reconstruction and reidentification attacks against public datasets, including attacks on Massachusetts Group Insurance Commission statistics (8), the Personal Genome project (9), the Netflix Prize dataset (10), the Airloak Challenge (11), and, most recently, the 2010 and 2020 Decennial Censuses (12–15).

simulated findings still support the original claims at significance level $\alpha = 0.1$; around 9% do not. At the scale of stronger privacy budgets ($\epsilon = 0.1$; lower ϵ entails stronger privacy and more infused noise), around 77% of findings still support the original claims. Notably, these rates of epistemic parity are generally higher than rates of parity in traditional large-scale replication studies in economics and psychology (28–30).

We analyze the antecedents of disparities due to added noise, quantifying the impacts of privacy budget, choice of noise-generating 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.

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 (31, 32). We find that even modest amounts of existing data error can alter findings at higher rates than typical amounts of privacy noise. Differentially private mechanisms do exacerbate these existing disparities—but the marginal impacts of privacy protection are small.

Our research builds on several studies examining the trade-offs between privacy and usability in public statistics, particularly in the context of the US census (32–37). Existing studies on research impacts examine synthetic microdata (eight studies replicated) (27) and DP linear regression (one simulation study) (38). These DP methods may be critical in future microdata products but are not yet widespread. These studies develop valuable benchmarks for future DP mechanism design, but do not provide larger-scale, systematic measures of current mechanisms suitable for contemporary policy debates. Our study examines, at scale, the predominant practice of infusing noise into aggregate statistics before public release, used in applications such as the 2020 Decennial Census (1) and the Social Science Facebook URLs dataset collaboration (2). Our method of analysis is also inspired by replication efforts in psychology and economics, particularly the Open Science Collaboration (28, 29) and the Many Labs replication project (39).

Recent US federal policy calls for the advancement of privacy-preserving data sharing across public and private sectors (19, 20), but a new Department of Commerce order specifically bans the use of noise infusion in future Census Bureau data products (40). The impacts of noise infusion policies can be widespread. The Integrated Public Use Microdata Series (IPUMS) alone, for example, lists over 27,000 academic publications based on its catalog of US 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 (22, 25, 43).

Our study provides an estimation of some of the key trade-offs involved: within the types of studies we replicated, with the simple privacy mechanisms we used, and at levels of protection typically used in industry, noise introduced for privacy was less likely to eliminate or reverse findings than the variation typically observed in robustness checks or replication studies. Moreover, the impacts of these privacy protections on findings were less severe than the impacts of simulating existing sources of data error in social statistics. These results could inform policymakers and data curators balancing privacy and utility in the next generation of privacy-preserving technologies

and inspire more robust, noise-aware approaches to data analysis.

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 Journals*, and several other journals (SI Appendix, Fig. S3). We searched for studies from the last two decades involving aggregate statistics about individuals or households. We only considered studies with findings based on regressions about individuals or households computed from aggregated data, rather than analyses of solely individual-level microdata. Of 166 studies that met these criteria, we successfully reproduced 93.

For each article, we used the abstract and introduction to identify the key empirical claims made by the author. We then identified the statistical estimate or estimates supporting those claims—in our sample, all regression coefficients—and modified the authors' code to extract those key numerical estimates and related statistics (SE, *t*-statistic, degrees of freedom, and so forth). For most studies, we reproduced 2 to 4 estimates (Table 1). Most estimates we reproduced (269 of 303) substantiate a key empirical claim mentioned in the article's abstract. Before noise injection, all 303 estimates we reproduced exactly matched the original estimates as published.

To simulate differentially private data release, we injected noise into most statistics about personal data used to produce the key findings in our sample, then ran the same analysis again using the noisy statistics. This approach models the likely scenario where authors do not adjust their analysis to account for the additional noise, comparable to treating it as unknown measurement error (46).[†] While burgeoning research proposes multiple imputation and data cleaning methods for noise-adjusted analysis of differentially private data (47–49), these efforts are nascent, currently difficult to apply universally, and not widely adopted. Indeed, as recently as 2024, the majority of economists were not even familiar with differential privacy (50), and data curators' noise infusion methods are not always transparent enough to easily account for in analysis (46).

We compared two common noise injection mechanisms: 1) pure differential privacy with the Laplace mechanism (18), a simple randomized mechanism for releasing statistics; and 2) zero-concentrated differential privacy (zCDP) (51, 52) with the Gaussian mechanism, designed to improve utility with a more relaxed definition of DP popular in industry (3, 4). 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*). A mechanism that satisfies pure ϵ -(pure) DP also satisfies $\rho = \left(\epsilon \frac{\exp(\epsilon)-1}{\exp(\epsilon)+1} \right)$ -zCDP (53).

[†] Comparable to classifying the added noise as unknown additive measurement error, this “plug-in” approach may not preserve Type I error rates and other frequentist properties (46).

Table 1. Descriptive statistics

	Obs.	Mean (SD)	Median (Q1, Q3)	Range
Studies	93			
Years since publication	93	7.52 (4.67)	7.00 (4.00, 11.0)	0 to 19.0
Num. results replicated	93	3.26 (1.86)	3.00 (2.00, 4.00)	1.00 to 10.0
Citations (Semantic Scholar)	92	186.0 (515.0)	57.5 (11.8, 124.0)	0 to 3970.0
Results	303			
Model deg. of freedom (num. predictors – 1)	292	60.3 (131.0)	20.0 (9.00, 40.5)	1.00 to 823.0
Num. vars. about people	303	8.20 (12.6)	5.00 (3.00, 10.5)	1.00 to 110.0
Proportion of personal vars. privatized	303	0.842 (0.238)	1.00 (0.775, 1.00)	0.167 to 1.00
Avg. sample unit population size	279	1.49e+6 (6.83e+6)	1.63e+4 (980., 1.04e+5)	0.171 to 5.37e+7
Answers primary or coprimary RQ	248			
Supports claim mentioned in abstract	269			
Includes individual-level microdata (not noised)	24			
Personal data variables	1026			
Successfully privacy protected	859			
<i>Role in regression</i>				
Control var.	707			
Dependent var.	152			
Primary independent var.	67			
Instrumental var.	19			
Instrumented var.	27			
Subsetting var.	8			
Weighting var.	38			
<i>Type of statistic</i>				
Ratio	32			
Booleans/categorical	56			
Count, log count	146			
Mean, log mean	517			
Median	22			
More complex query	242			

Epistemic Parity

Following ref. 27, we test for epistemic parity: the principle that empirical claims should be replicable with a differentially private dataset. A claim is an epistemic assertion based on one or more findings (27, 54) and a finding is a comparison between a statistical estimate 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. (55) 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 CIs within this region at a given confidence level $1 - \alpha$. 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. We except quantitative claims not originally based on statistical significance (e.g., a claim that an estimate falls between 0.1 and 0.2 despite its original CI exceeding those bounds) (see SI Appendix, Fig. S1 for examples).

We also test other common metrics of epistemic parity that do not depend on our interpretation of the authors’ claims: sign parity and sign reversal (whether the direction of effect matches or does not match the original at significance level α) and CI coverage (whether the CI contains the original estimate at confidence level $1 - \alpha$).

Impacts of Noise for Differential Privacy

The left column of Fig. 1 presents the impacts of noise injected for differential privacy on several standardized metrics of epistemic parity developed in prior replication studies (27, 29, 38). Additional metrics at additional significance levels (effective sample size and absolute difference) are depicted in SI Appendix, Figs. S8 and S9.

Though differential privacy advocates often prefer $\epsilon < 1$ (56), 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 COVID-19 Community Mobility Reports used $\epsilon = 2.64$ per day, Facebook’s URLs dataset was released with ($\epsilon = 1.453$, $\delta = 10^{-5}$)-DP (using approximate DP, a relaxed definition of DP). The 2020 US Decennial Census Demographic and Housing Characteristics tables used $\rho = 4.96$ per decade, equivalent to ($\epsilon = 26.34$, $\delta = 10^{-10}$)-DP (3–5, 57).

At $\epsilon = 10$ with the $\epsilon \frac{\exp(\epsilon)-1}{\exp(\epsilon)+1}$ -zCDP Gaussian mechanism (a relatively weak privacy guarantee), we observe strict epistemic parity at significance level $\alpha = 0.1$ in around 95.8% of simulated findings (over 10 replicates per finding $\times$ 303 findings = 3,030 simulated findings)—simulations where the findings, 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 findings nullified by noise. The rate of originally insignificant findings found to be significant

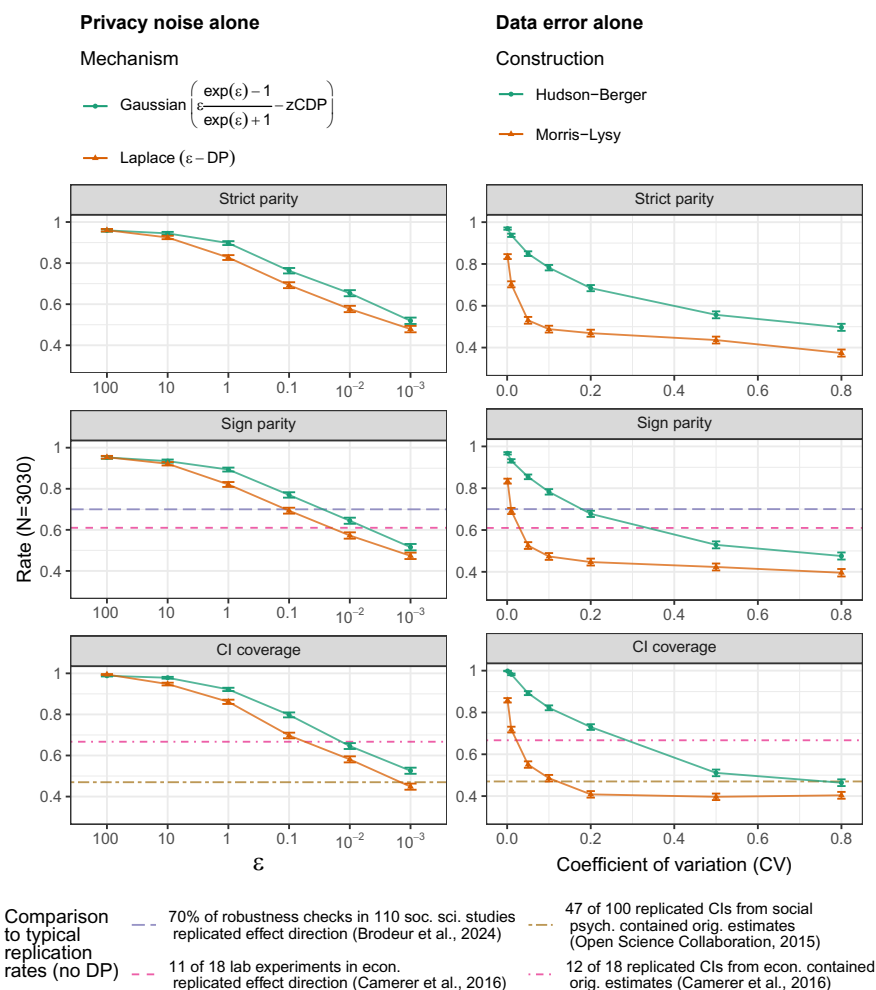


Fig. 1. Impacts of DP mechanisms alone (*Left*) and simulated added data error alone (*Right*) evaluated at significance level $\alpha = 0.05$, averaged over 303 findings times 10 simulations. Error bars depict 90% CI. Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators commonly use $0.1 \leq \epsilon \leq 10$.

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.

For some statistics in our sample (16.3%, affecting 134 of 303 findings), the original replication package did not include enough statistical information to simulate DP (see *Materials and Methods* for details). We did not add noise to those variables—similar to a scenario where a study combines statistics from some public data curators that use DP and some that do not. If we consider only the fully differentially private findings where privacy noise was added to *all* personal variables, rates of parity decrease slightly more sharply with ϵ : the rate of parity for fully privatized findings is 88.1% at $\epsilon = 1$ (vs. 90.9% for all findings) and 70.2% at $\epsilon = 0.1$ (vs. 77.4% for all findings). See *SI Appendix, Figs. S21–S23* for comparison.

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. (30) 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).[‡] Similarly, a replication of 18 laboratory experiments from economics (28) 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 Open Science Collaboration project in social psychology found that only 47 of 100 replicated 95% CI contained their original effect sizes (29); 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 (50), 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. In our sample, sign disparity (similar to significance mismatch) at $\alpha = 0.05$ reaches 10.7% at $\epsilon = 1$ with the zCDP Gaussian mechanism. Around 60% of economics researchers said they were willing to accept

[‡] 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 (58)—equivalent to 69% sign parity, similar to the rate of successful robustness checks (30).

Table 2. Effects on strict epistemic parity, averaged over 10 replicates

	(1)	(2)	(3)
(Intercept)	0.082 (0.254)	−0.038 (0.230)	0.664*** (0.104)
Log epsilon	0.089*** (0.007)	0.061*** (0.012)	0.018 (0.015)
Gaussian mech. (zCDP)	0.035*** (0.004)	0.035*** (0.004)	0.036*** (0.005)
Log epsilon x Gaussian mech. (zCDP)	−0.004** (0.001)	−0.004** (0.001)	−0.004* (0.001)
Log sum (Sensitivity/mean)	−0.033 (0.026)	−0.038 (0.021)	−0.034*** (0.010)
Original effect size		0.145 (0.124)	0.224** (0.084)
Original effect size × log epsilon		0.055* (0.022)	0.053* (0.021)
Claim: insignificant		0.246** (0.094)	0.261*** (0.059)
Claim: non-zero upper/lower bound		0.093 (0.062)	0.112* (0.046)
Log regression sample size			−0.019 (0.015)
Noised ind./treatment var.			−0.036 (0.043)
Noised ind./treatment var. × Log epsilon			0.005 (0.016)
Noised dep./outcome var.			−0.172*** (0.039)
Noised dep./outcome var. × Log epsilon			0.070*** (0.012)
Avg. sample unit population size			0.000 (0.000)
Study FE	Yes	Yes	No
Implementation controls	Yes	Yes	Yes
Query type controls	Yes	Yes	Yes
N. obs.	29,400	28,322	25,774
R squared	0.545	0.574	0.509
F statistic	340.210	369.724	1212.407

*** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$. Robust SEs are clustered by study. Controls include number and type of component queries, number of variables noised, and number of predictors in the regression. Dummy for “Claim: sig. positive/negative” 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).

estimates switching sign (similar to sign reversal) 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% ($\alpha = 0.05$; *SI Appendix, Fig. S9*).

Antecedents of Epistemic Disparity

Table 2 summarizes a linear model of the effect of privacy and sensitivity via noise mechanism (Column 1), the type of epistemic claims made (Column 2), and sample unit size and modeling choices made during regression analysis (Column 3) on epistemic parity. We regress over the average strict epistemic parity at $\alpha = 0.1$ for each finding ($N = 303$) 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 $[10^{-5}, 10^3]$.

Privacy and Sensitivity. 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, the maximum amount its output

could change when an individual’s data is added or removed (*Definition 2*). *SI Appendix, Table S2* demonstrates this model in isolation; the root mean squared deviation applied to each component query is almost completely explained by ϵ , the sensitivity of the query, and the choice of mechanism. We add each of these factors to the findings-level regression in Column 1, aggregating to the findings level by totaling the sensitivity of each component query divided by its original value. The total noise added to each statistic used in analysis depends on not only the noise added to its component queries but also 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 (e.g., mean).

As expected, epistemic parity decreases with sensitivity and increases with ϵ . Averaged across the values of ϵ we tested, a 10-fold increase in ϵ corresponds to a 8.9 percentage point increase in the rate of strict epistemic parity, all else equal (Column 1, “Log epsilon”) (see *SI Appendix, Fig. S7* for precise effects of each of the 50 values of ϵ tested). This estimate is robust to the introduction of other controls (Columns 2 and 3). Using zCDP with the Gaussian mechanism corresponds to a 3.5 – 0.4 log ϵ percentage point increase over the Laplace mechanism [Column 1, “Log epsilon × Gaussian mech. (zCDP)”. The Gaussian mechanism performs slightly better than the Laplace mechanism for $\epsilon < 10$ (*Fig. 1*, left column).

Epistemic Claims. 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. *Fig. 2* depicts this finding graphically with results from the ϵ -DP Laplace mechanism and the standardized effect size correlation coefficient r used by Open Science Collaboration (29). In our regression analysis, a 0.1 decrease in the original effect size corresponds to a $1.45 + 0.55 \log \epsilon$ percentage point decrease in the rate of epistemic parity, other factors held equal (Column 2, “Original effect size × Log epsilon”).

This finding mirrors robustness research in social science. Brodeur et al. (30), for example, find that half of robustness checks of point estimates significant at $0.05 < P < 0.1$ succeed,

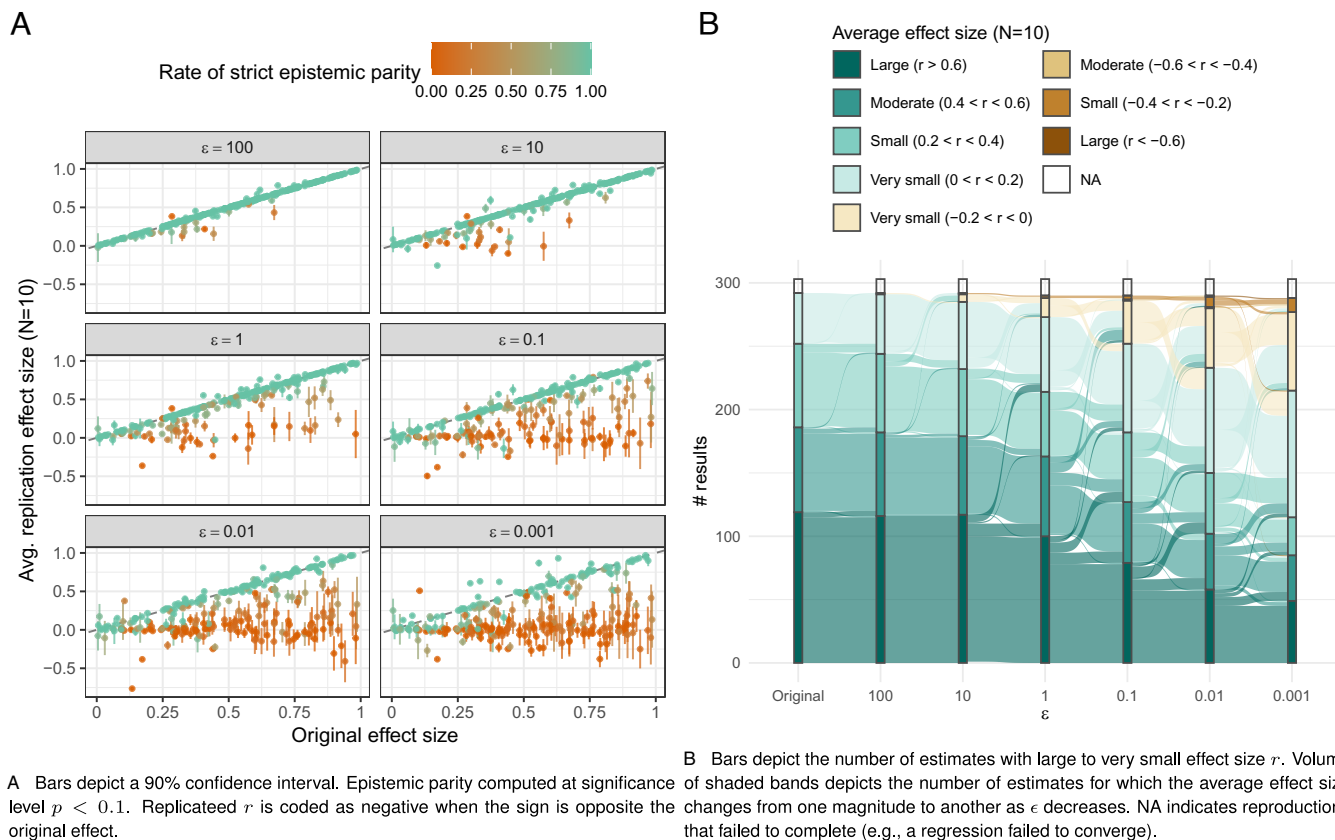


Fig. 2. Change in average standardized effect size (r), averaged over 10 simulations per estimate, as a result of applying the Laplace mechanism with decreasing privacy parameter ϵ . Lower values of the privacy parameter ϵ provide stronger privacy protection but require more injected noise; curators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

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 SEs inflate in the presence of measurement error (59, 60). Attenuation is clearly visible in the downward shift in the distribution of effect sizes as ϵ decreases (SI Appendix, Figs. S8 and S10). 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 (Fig. 2A). In robustness checks of 110 social science studies, Brodeur et al. (30) observe significance levels 77% of the original; in replication of 18 economics laboratory experiments, Camerer et al. (28) 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 SEs increase (61). 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 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 (Fig. 2A). For $\epsilon > 1$, epistemic disparities come almost exclusively from findings with initially lower effect sizes (Fig. 2A). Most disparities occur when the original claim requires a coefficient with a certain sign or within a certain interval (SI Appendix, Fig. S13). Null findings (of statistical insignificance), on the other hand, correspond to 24.6 percentage points higher epistemic parity than standard direction of effect claims, all else equal (Column 2,

“Claim: insignificant”). The CI coverage and difference in effect size, which do not depend on the authors’ claims about their findings, are less sensitive to small amounts of noise. (The rate of 90% CI coverage is 99.7% with the Laplace mechanism at $\epsilon = 1000$.)

Analysis and Sample Size. Column 3 adds characteristics of the original regression analysis used to produce the findings. (In this specification, we drop study fixed effects because some studies use the same regression model for all findings.) Especially important is the size of the sample unit used to produce each statistic relative to the sensitivity of the statistic. SI Appendix, Figs. S16 and S17 shows that rates of epistemic parity are generally higher for statistics covering larger geographies or time periods (e.g., state-level vs. university-level, or annual vs. daily). These larger area statistics usually have the same sensitivity as smaller areas (e.g., a simple population count with sensitivity 1), and thus less noise proportional to their size. In Column 3, we control for both the average sample unit size if it is available in the dataset (e.g., average county population for a regression over counties) and the sum of the ratios of sensitivity to each original component query used in the analysis. The effect of average sample unit size alone is negligible (SI Appendix, Fig. S11), but a 10-fold increase in the sum of normalized sensitivity corresponds to a 3.8 percentage point decrease in the rate of strict parity (Column 3, “Log sum (Sensitivity/mean)”; see also SI Appendix, Fig. S12).

Query sensitivity helps to explain a few extreme disparities we observe even when the privacy guarantee is extremely weak (Fig. 2A). For example, at $\epsilon = 1,000$, there are still disparities in a rare 0.8% of cases (Laplace mechanism), all from just seven findings. 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 (55) or monthly neighborhood crime rates (62). 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 than the ones we simulate here (63). Removing noise for privacy from only these high-sensitivity queries eliminates these high- ϵ disparities in all cases.

Attenuation appears especially acute when the outcome or dependent variable is privatized (*SI Appendix, Fig. S15*). Adding noise to the outcome variable in a regression leads to $17.2 - 7.0 \log \epsilon$ percentage points lower strict epistemic parity, other factors held equal (Column 3, “Noised dep./outcome var.,” “Noised dep./outcome var. $\times$ Log epsilon”). The privacy-utility curve for findings utilizing instrumental variables (IV) regressions is noticeably steeper than for other regressions; IV estimates tend to result in lower rates of epistemic parity (*SI Appendix, Fig. S14*).

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; *SI Appendix, Fig. S18*.)

We observe similar trends in the rate of epistemic parity across types of personal data (*SI Appendix, Fig. S19*). 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 (64). 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 study privacy utility rates by type of data source: administrative, census, 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 (*SI Appendix, Fig. S20*). 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 (65–67). Work is still ongoing to explore these considerations and develop optimal DP-compatible design strategies for surveys (66–68) and small-area estimation (69).

Impacts of Existing Data Error

Social statistics contain many different kinds of existing data error, including measurement error, sampling error, and model error (31, 32, 70). To compare the effects of DP mechanisms to

the possible impacts of existing data error—modeled here naively as zero-bias sampling error (70)—we simulate possible alternative versions of the dataset by drawing counterfactual observations from a normal distribution with variance proportional to the original statistic. We test two admissible constructions using shrinkage estimation from multilevel Bayesian modeling (71): the Hudson–Berger construction (72), which generally shrinks estimates more when sampling variance is high, and the Morris–Lysy construction, which shrinks estimates more when sampling variance is low (73).

We find that larger epistemic disparities occur if statistics contain even modest amounts of data error (Fig. 1, right column). US Census Bureau American Community Survey guidance describes a 1.5% coefficient of variation—the ratio of SD to the original statistic—as “very small” and “very reliable” for aggregated count data (74). Simulating normally distributed sampling 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. A regression analysis depicted in *SI Appendix, Table S4* confirms some of the same antecedents from our privacy experiments—larger original effect sizes and null findings (originally insignificant results) are associated with lower disparities.

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 due to sampling error of less than 30% (70); more complicated data products, such as the school district-level Small Area Income and Poverty Estimates, are published with coefficients as high as 67% (75).

Researchers may 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 (76). Similar overconfidence may extend to robustness to existing data error.

Given the potentially large effects of data 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. Fig. 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 2.9 percentage points.

Discussion

Our sample is only representative of reproducible social science studies involving regressions over aggregate statistics about individuals or households. We sample studies from repositories of replication packages and require that the replication packages work and contain sufficient information to implement our

Privacy noise alone

Mechanism

$$\text{Gaussian} \left(\frac{\exp(\epsilon) - 1}{\exp(\epsilon) + 1} - \text{zCDP} \right)$$

Laplace ($\epsilon - \text{DP}$)

Data error (Hudson-Berger) + privacy noise (Gaussian)

Data error CV

0.001 0.1 0.5

0.01 0.2 0.8

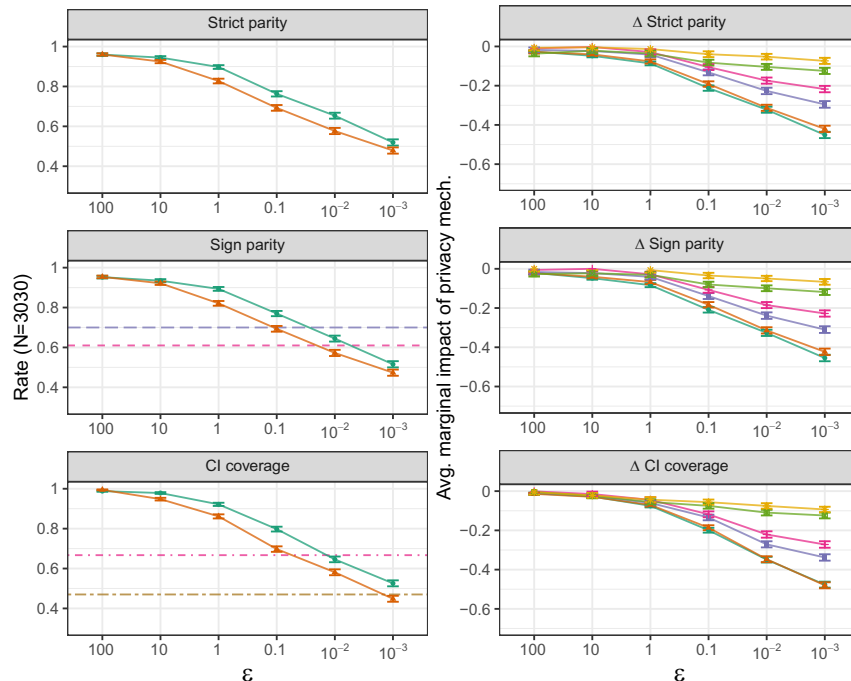


Fig. 3. Impacts of DP mechanisms alone (*Left*) and the marginal impact of the zCDP Gaussian mechanism *after* various levels of data error (*Right*), evaluated at significance level $\alpha = 0.05$, averaged over all findings and 10 simulations. Error bars depict 90% CI. Lower values of the privacy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

Comparison to typical replication rates (no DP) — 70% of robustness checks in 110 soc. sci. studies replicated effect direction (Brodeur et al., 2024) — 47 of 100 replicated CIs from social psych. contained orig. estimates (Open Science Collaboration, 2015)
 — 11 of 18 lab experiments in econ. replicated effect direction (Camerer et al., 2016) — 12 of 18 replicated CIs from econ. contained orig. estimates (Camerer et al., 2016)

privacy simulations; less well documented studies may exhibit higher disparities. Also, the keywords used in our search (such as “county,” “school,” “neighborhood,” or “organization”) aim for representation but favor geographic regions and organizations, which may exclude other types of aggregation (e.g., temporal). The studies we reproduced span a wide range of sample unit sizes (*SI Appendix, Fig. S11*), data domains (*SI Appendix, Fig. S5*), and academic disciplines (*SI Appendix, Fig. S3*), but studies involving microdata, aggregation levels not included in our list of keywords (*SI Appendix, section A*), or other approaches to privacy protection may exhibit different effects.

With these caveats, 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 (28, 30). 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 nontrivial. Numerous studies have shown that replications of findings from psychology and economics do not support the original conclusions at similar or higher rates than what

we observe (28, 29, 39, 58). Other studies similarly show how findings can be sensitive to small changes in the data; Broderick et al. (77), for example, find that the findings 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 (78): 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 (1), further reducing disparities. There are still some ways our results may underestimate the impact of differential privacy: in the real world, these statistics have to share privacy budget with other statistics within large data products; microdata may be privacy protected as well as aggregate statistics; 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, including microdata analysis, machine learning, and other methods where noise infusion may be more or less impactful.

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 (31, 32, 46)—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 (61). There are long-established methods for calibrating regression estimates to account for additive error (60, 79). 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 (48, 49). Agarwal et al. (47), for example, propose a method that could recover the main estimates of Autor et al. (80) (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 methods to account for injected noise when Meta released its differentially private Facebook URLs Dataset (48, 81). Future studies could explore the impacts of replicating existing studies with noise-aware methods throughout data exploration and analysis.

While adopting new methods is costly, there can be benefits to scientific practice. For example, Dwork et al. (82) and Echenique and He (83) 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 toward practices that are both more private and more robust (84).

Materials and Methods

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 include articles from journals with standardized data accessibility policies, including American Economic Association (AEA) journals, the *Review of Economics and Statistics*, and several other journals (SI Appendix, Fig. S3). 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. SI Appendix, section A 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 reproducible contemporary research involving statistics about individuals aggregated at common sampling levels, we restricted the sample to articles published in the last 19 years that use personal data collected within the last 70 y (similar to the US Census Bureau's policy of keeping data confidential for 72 y after collection). We also excluded three studies with code primarily written MATLAB, which was not supported by our computing infrastructure. The final sample consisted primarily of Stata and R scripts. We attempted to

reproduce 166 studies that met these criteria,[‡] correcting the authors' code only as much as necessary to reproduce their estimates without error. We recorded the expected values of key estimates reported in the article and in every run, and we automatically checked that those values matched the estimates produced by our copy of the authors' code. All of our reproductions matched the authors' original estimates to the same level of numerical precision reported in their published article. We were unable to reproduce 73 studies, either because the replication package did not include the datasets necessary to reproduce the main estimates, because the datasets did not include enough information to implement differential privacy, or, less often, because of bugs or discrepancies in the replication packages.

Our sample includes 93 successfully replicated studies, 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 supporting those claims—in our sample, all regression coefficients—and modified the authors' code to extract those key numerical estimates and related statistics (SE, *t*-statistic, degrees of freedom, etc.). To standardize our analysis, we did not consider visual findings (multiple estimates represented by visual relation in a figure) (27). Visual findings substantiated key claims in only a small handful of the studies we found. For most studies, we reproduced 2 to 4 key findings (Table 1). Most estimates we reproduced (269 of 303) substantiate a key empirical claim mentioned in the article's abstract (Table 1).

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. $H_A : \beta \neq 0$.

CI coverage at α : whether the original estimate falls within the CI of the counterfactual estimate at confidence level $1 - \alpha$. 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 (87) and *P*-values can vary widely over replications of the same test (88). CI may give more reliable information about replicability.

Bias in standardized effect size: the absolute difference between the counterfactual effect size and the original effect size. We follow SI Appendix, section A3 (29) in converting each estimate to a common effect size statistic, the correlation coefficient *r* computed from the *t*-statistic and residual degrees of freedom *v* for each estimate:

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

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

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 findings in our sample.

We compare two approaches: 1) pure differential privacy with the Laplace mechanism (18), a simple randomized mechanism for releasing statistics; and 2) zero-concentrated differential privacy (zCDP) (51, 52) with the Gaussian mechanism, a relaxation of differential privacy with improved utility. 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 (63).

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 (Definition 1).

[‡]The terms “repeatability”, “reproducibility”, and “replicability” are often interchanged (85). Here, we use the term “reproduce” to mean achieving the same findings with the same data and methods and “replicate” to mean achieving similar findings using the same methods but different (noisy) data (86).

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

Definition 1 [18] Let $M : \mathcal{D}^n \rightarrow \mathcal{Y}$ be a randomized algorithm. M is ϵ -(pure) differentially private if for all pairs of datasets $D, D' \in \mathcal{D}^n$ which differ in only one entry and all possible outputs $T \subseteq \mathcal{Y}$:

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

Pure DP with the Laplace Mechanism. To simulate releasing $F(D)_{k \times v}$ with ϵ -(pure) differential privacy, we apply the Laplace mechanism to each query $f_j(D)$ (Definition 3). We rely on basic composition (89, Corollary 3.15) to release all the queries $F(D)_{k \times v}$ with ϵ -DP: we use $\frac{\epsilon}{v}$ -DP mechanisms $M_j^{\text{Lap}}(D)$ for each query f_j such that $M^{\text{Lap}}(D) = (M_1^{\text{Lap}}(D), M_2^{\text{Lap}}(D), \dots, M_v^{\text{Lap}}(D))$ is ϵ -(pure) DP.

This is an “input perturbation” approach to regression modeling (90). 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 (1, 48)]. 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 (90), but this setup is less common and requires more effort from both data curators and researchers (37).

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 a single 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) (4). With that assumption, $\Delta f_j(D) = \Delta f_j(d_i) \forall d_i \in D$ (equivalent to parallel composition).

Definition 2 [89] The l_p -sensitivity of a function $f : \mathcal{X}^n \rightarrow \mathbb{R}^k$ is

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

where D, D' are datasets which differ by exactly one added or removed entry (resulting in “unbounded” DP (91), which does not treat the dataset size as fixed).

Definition 3 [89] Let $f : \mathcal{D}^n \rightarrow \mathbb{R}^k$. The Laplace mechanism is defined as

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

where the Y_i are independent random variables drawn from the Laplace distribution $Y_i \sim \text{Lap}\left(\frac{\Delta f}{\epsilon}\right)$ (Definition 1), and Δf is the l_1 -sensitivity of the query function f (Definition 2).

zCDP with the Gaussian Mechanism. Zero-concentrated differential privacy (zCDP; Definition. 4) is a relaxation of pure differential privacy that provides better accuracy (51, 52). The Gaussian mechanism (Definition 5) satisfies zCDP. This mechanism was used to release statistics from the 2020 Decennial Census (1).

zCDP uses a different privacy parameter, ρ . A mechanism that satisfies pure DP also satisfies zCDP; specifically, if M satisfies ϵ -DP, then M satisfies $\left(\epsilon \frac{\exp(\epsilon)-1}{\exp(\epsilon)+1}\right)$ -zCDP (52, 53). We use this bound for comparison between the two mechanisms hereafter. zCDP has the same basic composition property as

pure DP (52, Lemma 1.7), so we again apply $\frac{\rho}{v}$ -zCDP mechanisms $M_j^{\text{Gauss}}(D)$ such that $M^{\text{Gauss}}(D) = (M_1^{\text{Gauss}}(D), M_2^{\text{Gauss}}(D), \dots, M_v^{\text{Gauss}}(D))$ is ρ -zCDP.

Definition 4 [52] Let $M : \mathcal{D}^n \rightarrow \mathcal{Y}$ be a randomized algorithm. M is ρ -zCDP if for all pairs of datasets $D, D' \in \mathcal{D}^n$ 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 $M(D')$ (SI Appendix, Definition 1).

Definition 5 [52] Let $f : \mathcal{D}^n \rightarrow \mathbb{R}^k$. The Gaussian mechanism is defined as

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

where the Z_i are independent random variables drawn from the Gaussian distribution $Z_i \sim \mathcal{N}\left(0, \frac{(\Delta f)^2}{2\rho}\right)$, and Δf is the l_2 -sensitivity of the query function f (Definition 2).

Sensitivity and Postprocessing. For each study in our sample, we identified all the statistics about personal data used to produce the key findings. For each statistic, we defined the global sensitivity of the query used to produce each individual statistic (Definition 2). When the global sensitivity of a query was unbounded (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. We included these pre- and postprocessing computations even when no noise was injected, automatically checking that our reconstructions preserved the original data and findings in each run. Some variables are used as regression weights and must not be negative; for those variables, we clipped the noisy values to zero. Similarly, some weights must be integers; these we round to match the original dataset. Differential privacy is preserved through these postprocessing steps (89, 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 aggregate statistics used in our sample, the original authors’ replication package did not provide enough information to define sensitivity directly and reconstruct the statistical query (SI Appendix, Figs. S4 and S5). 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 age is clipped at 120 y maximum).[#] We did not apply any privacy protection to the remaining statistics (16.3% of statistics in our sample, affecting 134 of 303 findings), either because sensitivity was unbounded (e.g., for ratios) and the component queries were not included in the dataset [e.g., for unemployment rates, neither the number of employed nor the size of the labor force] or because we could not reverse engineer the statistic using the available code and data (e.g., an intermediate model-based prediction of education funding based on race (94)). See SI Appendix, Figs. S21–S23 for comparison.
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., refs. 66 and 67).

[#]For three continuous component queries—affecting just seven findings across two studies (Indian household consumption in ref. 92 and weekly wages in ref. 93)—we did not have enough context to assume a reasonable data-independent bound. We assumed that the data curator clips these queries at three SDs above the average value of the aggregate statistics—a data-dependent choice that technically violates the DP guarantee, but still offers a realistic simulation of noise infusion. Our main results are nearly identical when these seven findings are removed (SI Appendix, Fig. S24). Even removing the 86 findings across 31 that involve queries with sensitivity greater than one (including all data-independent bounds) only slightly raises rates of epistemic parity (SI Appendix, Fig. S25).

- Statistics are likely to be released as part of larger datasets, and data curators may have to split privacy loss budgets among 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 be injected in practice.
- In our mechanisms, each statistic receives an equal share of the privacy budget. In real-world deployments such as the US Decennial Census, data curators may tune budgets more carefully to prioritize certain critical statistics (12), adding or reducing noise depending on policy priorities.
- Six of the studies in our sample (24 findings) include individual-level microdata in addition to aggregate statistics. Since this study only examines the effect of privatizing aggregate statistics, we did not add noise to the microdata—simulating the likely scenario where researchers combine privatized public statistics with confidential, unperturbed microdata.
- 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 preprocessing 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.
- 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.

SI Appendix, section B provides a more detailed, step-by-step description of our replication and noise infusion protocol with examples from our sample.

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 set sampling variance $v_{ij} = (c_{ij}|x_{ij}|^{1-b})^2$ in proportion to the magnitude of the observed statistic $|x_{ij}|$. If the original statistic is zero for a given sampling unit d_i , we impute the smallest nonzero value across sampling units $\min_{i \in [k]} f_j(d_i)$. The parameter b models the rate at which the SD $c_{ij}x_{ij}$ 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

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

is inadmissible for the true population statistics μ (95). Cui et al. (71) suggest two admissible constructions using shrinkage estimation from multilevel empirical

Bayesian modeling. Following Cui et al. (71), we simulate replicates with the general form

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

where β_j, B_{ij} are functions of x_{ij} and v_{ij} (71, 73). This method adjusts the counterfactual estimate to account for a baseline β_j with variance reduced by 100 B_{ij} %. The Hudson–Berger construction (72, 96) 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 (73) 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_{ij} , $\hat{B}_j^H = (k-3)/(k-1)\hat{\sigma}^2$, 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_{ij} .

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 inadmissible no-shrinkage construction (Eq. 2) for comparison.

Data, Materials, and Software Availability. All data and code that support the findings of this study can be accessed at <https://github.com/ryansteed/dp-reproducibility> (97). Due to size constraints, original replication datasets are not included, but we will provide instructions and links to obtain the datasets from the original sources. The replication codebases and our modifications are included.

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. This material is based upon work supported by the National Science Foundation under Grant No. 231919.

Author affiliations: ^aHeinz College of Information Systems and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213; and ^bSloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139

- J. M. Abowd et al., The 2020 census disclosure avoidance system TopDown algorithm. *Harv. Data Sci. Rev.*, 10.1162/99608f92.529e3cb9, (2022).
- G. King, N. Persily, Unprecedented Facebook URLs dataset now available for academic research through social science one. *Social Science One* (2020). <https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one>. Accessed 6 April 2023.
- D. Desfontaines, A list of real-world uses of differential privacy. *Ted Is Writing Things* (2021). <https://desfontain.es/privacy/real-world-differential-privacy.html>. Accessed 16 May 2022.
- P. Nanayakkara, E. Ghazi, S. Vadhan, Practitioners' Perspectives on a Differential Privacy Deployment Registry. *arXiv [Preprint]* (2025). <http://arxiv.org/abs/2509.13509> (Accessed 7 October 2025).
- M. Khavkin, E. Toch, Differential privacy configurations in the real world: A comparative analysis. *IEEE Trans. Knowl. Data Eng.* **37**, 6334–6351 (2025).
- J. M. Abowd, M. B. Hawes, Confidentiality protection in the 2020 US Census of population and housing. *Annu. Rev. Stat. Appl.* **10**, 119–144 (2023).
- A. Narayanan, V. Shmatikov, Myths and fallacies of "Personally Identifiable Information". *Commun. ACM* **53**, 24–26 (2010).
- L. Sweeney, K-anonymity: A model for protecting privacy. *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.* **10**, 557–570 (2002).
- L. Sweeney, A. Abu, J. Winn, Identifying participants in the Personal Genome Project by name (A re-identification experiment). *arXiv [Preprint]* (2013). <http://arxiv.org/abs/1304.7605> (Accessed 7 August 2024).
- A. Narayanan, V. Shmatikov, How To break anonymity of the Netflix Prize Dataset (2007).
- A. Cohen, K. Nissim, Linear program reconstruction in practice. *J. Priv. Confidentiality* **10** (2020).
- J. M. Abowd et al., A Simulated Reconstruction and Reidentification Attack on the 2010 U.S. Census. *Harv. Data Sci. Rev.* **7** (2025).
- T. Dick et al., Confidence-ranked reconstruction of census microdata from published statistics. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2218605120 (2023).
- R. Steed, D. Qing, S. Wu, Quantifying privacy risks of public statistics to residents of subsidized housing. *Harv. Data Sci. Rev.*, 10.1162/99608f92.39d8bfa4 (2025).
- A. Flaxman, O. Keyes, The risk of linked census data to transgender youth: A simulation study. *J. Priv. Confidentiality* **15** (2025).
- D. Boyd, J. Sarathy, Differential perspectives: Epistemic disconnects surrounding the U.S. Census bureau's use of differential privacy. *Harv. Data Sci. Rev.*, 10.1162/99608f92.66882f0e (2022).
- I. Dinur, K. Nissim, "Revealing information while preserving privacy" in *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems-PODS '03* (ACM Press, San Diego, CA, 2003), pp. 202–210.
- C. Dwork, F. McSherry, K. Nissim, A. Smith, "Calibrating noise to sensitivity in private data analysis" in *Proceedings of the Third Conference on Theory of Cryptography, TCC'06* (Springer-Verlag, Berlin/Heidelberg, Germany, 2006), pp. 265–284.
- J. R. Biden, Executive order on the safe, secure, and trustworthy development and use of artificial intelligence (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/>. Accessed 31 October 2023.
- National Science and Technology Council, "National strategy to advance privacy-preserving data sharing and analytics" (Tech. Rep., Executive Office of the President, 2023).
- S. Ruggles, C. Fitch, D. Magnuson, J. Schroeder, Differential privacy and census data: Implications for social and economic research. *AEA Pap. Proc.* **109**, 403–408 (2019).
- V. J. Hotz et al., Balancing data privacy and usability in the federal statistical system. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2104906119 (2022).
- M. Leanon, McEachin casts doubt on Census Bureau's plans for population count data. *Richmond Times-Dispatch* (2020).

24. S. A. Ochoa, T. A. Minnis, "Preliminary report: Impact of differential privacy & the 2020 census on Latinos, Asian Americans and redistricting" (Report, MALDEF & Asian Americans Advancing Justice, 2021).
25. D. Oberski, F. Kreuter, Differential privacy and social science: An urgent puzzle. *Harv. Data Sci. Rev.* **2** (2020).
26. J. H. Childs, USC Bureau, "Trust and credibility in the U.S. Federal Statistical System" in *Survey Methods: Insights from Field (SMIF)* (2019).
27. L. Rosenblatt, et al., Epistemic parity: Reproducibility as an evaluation metric for differential privacy. *Commun. ACM* **69**, 93–102 (2026).
28. C. F. Camerer et al., Evaluating replicability of laboratory experiments in economics. *Science* **351**, 1433–1436 (2016).
29. Open Science Collaboration, Estimating the reproducibility of psychological science. *Science* **349**, aac4716 (2015).
30. A. Brodeur et al., "Mass reproducibility and replicability: A new hope" (I4R Discussion Paper Series, Working Paper 107, 2024).
31. R. M. Groves, L. Lyberg, Total survey error: Past, present, and future. *Public Opin. Q.* **74**, 849–879 (2010).
32. R. Steed, T. Liu, Z. S. Wu, A. Acquisti, Policy impacts of statistical uncertainty and privacy. *Science* **377**, 928–931 (2022).
33. J. M. Abowd, I. M. Schmutte, An economic analysis of privacy protection and statistical accuracy as social choices. *Am. Econ. Rev.* **109**, 171–202 (2019).
34. C. T. Kenny et al., The use of differential privacy for census data and its impact on redistricting: The case of the 2020 U.S. Census. *Sci. Adv.* **7**, eabk3283 (2021).
35. D. Pujol 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., 2020), pp. 189–199.
36. Q. Brummet, E. Mulrow, K. Wolter, The effect of differentially private noise injection on sampling efficiency and funding allocations: Evidence from the 1940 census. *Harv. Data Sci. Rev.*, 10.1162/99608f92.a93d96fa (2022).
37. A. F. Barrientos, A. R. Williams, J. Snoko, C. M. Bowen, A feasibility study of differentially private summary statistics and regression analyses with evaluations on administrative and survey data. *J. Am. Stat. Assoc.* **119**, 52–65 (2024).
38. A. R. Williams, A. F. Barrientos, J. Snoko, C. M. Bowen, Benchmarking DP linear regression methods for statistical inference (2024). https://conference.nber.org/conf_papers/f194244.pdf. Accessed 17 July 2024.
39. R. A. Klein et al., Investigating variation in replicability. *Soc. Psychol.* **45**, 142–152 (2014).
40. G. M. Cook, Understanding the new disclosure avoidance policy: What it means for data, confidentiality and you (2026). <https://www.census.gov/newsroom/blogs/directofr/2026/08/understanding-the-new-disclosure-avoidance-policy.html>. Accessed 31 August 2026.
41. IPUMS, Bibliography (2024). https://bibliography.ipums.org/citations/results?search_terms%5Bmax_year_published%5D=2024&search_terms%5Bmin_year_published%5D=1960. Accessed 7 August 2024.
42. L. Penberthy, SEER marks 50 years of turning cancer data into discovery. *NIH Rec. LXXV* **12** (2023).
43. A. Acquisti, R. Steed, Learning to live with privacy-preserving analytics. *Commun. ACM* **66**, 24–27 (2023).
44. Inter-University Consortium for Political and Social Research, Find data (2025). <https://www.icpsr.umich.edu/web/pages/ICPSR/index.html>. Accessed 16 April 2025.
45. S. Kranz, Find economic articles with data (2024). <https://ejd.econ.mathematik.uni-ulm.de/>. Accessed 31 July 2024.
46. R. Gong, Transparent privacy is principled privacy. *Harv. Data Sci. Rev.*, 10.1162/99608f92.b5d3faaa (2022).
47. A. Agarwal, R. Singh, Causal inference with corrupted data: Measurement error, missing values, discretization, and differential privacy. arXiv [Preprint] (2024). <https://arxiv.org/abs/2107.02780> (Accessed 26 February 2024).
48. G. Evans, G. King, Statistically valid inferences from differentially private data releases, with application to the Facebook URLs dataset. *Polit. Anal.* **31**, 1–21 (2023).
49. M. Blackwell, J. Honaker, G. King, A unified approach to measurement error and missing data: Overview and applications. *Sociol. Methods Res.* **46**, 303–341 (2017).
50. A. R. Williams, J. Snoko, C. M. Bowen, A. F. Barrientos, Disclosing economists' privacy perspectives: A survey of American Economic Association Members' views on differential privacy and the usability of noise-infused data. *Harv. Data Sci. Rev.*, 10.1162/99608f92.a8b0371 (2024).
51. C. Dwork, G. N. Rothblum, Concentrated differential privacy. arXiv [Preprint] (2016). <https://arxiv.org/abs/1603.01887> (Accessed 10 March 2025).
52. M. Bun, T. Steinke, "Concentrated differential privacy: Simplifications, extensions, and lower bounds" in *Proceedings, Part I, of the 14th International Conference on Theory of Cryptography - Volume 9985* (Springer-Verlag, Berlin, Heidelberg, 2016), pp. 635–658.
53. T. Steinke, Tight RDP & zCDP bounds from pure DP (DifferentialPrivacy.org) (2024). <https://differentialprivacy.org/pdp-to-zcdp/>. Accessed 10 March 2025.
54. K. B. Cohen et al., "Three dimensions of reproducibility in natural language processing" in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, N. Calzolari et al., Eds. (European Language Resources Association (ELRA), Miyazaki, Japan, 2018).
55. B. Guidetti, P. Pereda, E. Severini, "Placebo Tests" for the impacts of air pollution on health: The challenge of limited health care infrastructure. *AEA Pap. Proc.* **111**, 371–375 (2021).
56. C. Dwork, N. Kohli, D. Mulligan, Differential privacy in practice: Expose your epsilons! *J. Priv. Confidentiality* **9** (2019).
57. US Census Bureau, Privacy-loss budget allocation 2023-04-03 (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. Accessed 28 July 2023.
58. R. Silberzahn et al., Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Adv. Methods Pract. Psychol. Sci.* **1**, 337–356 (2018).
59. J. Hausman, Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *J. Econ. Perspect.* **15**, 57–67 (2001).
60. R. J. Carroll, D. Ruppert, L. A. Stefanski, C. M. Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective* (Chapman and Hall/CRC, New York, NY, ed. 2, 2006).
61. E. Loken, A. Gelman, Measurement error and the replication crisis. *Science* **355**, 584–585 (2017).
62. C. Bellégo, J. Drouard, Fighting crime in lawless areas: Evidence from slums in Rio de Janeiro. *Am. Econ. J.: Econ. Policy* **16**, 124–159 (2024).
63. R. Cummings et al., Advancing differential privacy: Where we are now and future directions for real-world deployment. *Harv. Data Sci. Rev.* **6**, (2024).
64. A. R. Santos-Lozada, J. T. Howard, A. M. Verdery, How differential privacy will affect our understanding of health disparities in the United States. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 13405–13412 (2020).
65. M. Bun, J. Drechsler, M. Gaboardi, A. McMillan, J. Sarathy, Controlling privacy loss in sampling schemes: An analysis of stratified and cluster sampling. arXiv [Preprint] (2023). <http://arxiv.org/abs/2007.12674> (Accessed 18 April 2025).
66. J. Drechsler, J. Bailie, The complexities of differential privacy for survey data. arXiv [Preprint] (2024). <http://arxiv.org/abs/2408.07006> (Accessed 17 April 2025).
67. J. Seeman, Y. Si, J. P. Reiter, Toward Differentially Private Finite Population Estimation: An Approach Based on Survey Weight Regularization. *Harv. Data Sci. Rev.* **8**, 10.1162/99608f92.fdb2338d (2026).
68. U.S. Census Bureau, Disclosure avoidance protections for the American Community survey (2022). <https://www.census.gov/newsroom/blogs/random-samplings/2022/12/disclosure-avoidance-protections-acsc.html>. Accessed 18 April 2025.
69. J. Seeman, A. Slavkovic, M. Reimherr, "Private posterior inference consistent with public information: A case study in small area estimation from synthetic census data in privacy in statistical databases" in *Lecture Notes in Computer Science*, J. Domingo-Ferrer, K. Muralidhar, Eds. (Springer International Publishing, Cham, Switzerland, 2020), pp. 323–336.
70. U.S. Census Bureau, U.S. Census bureau statistical quality standards (2013). <https://www2.census.gov/about/policies/quality/quality-standards-jul2013.pdf>. Accessed 20 March 2025.
71. Y. Cui, R. Gong, J. Hannig, K. Hoffman, Technical comment on "Policy impacts of statistical uncertainty and privacy". *Science* **380**, ead9724 (2023).
72. H. M. Hudson, "Empirical Bayes estimation" (Tech. Rep., 1974).
73. C. N. Morris, M. Lysy, Shrinkage estimation in multilevel normal models. *Stat. Sci.* **27**, 115–134 (2012).
74. U.S. Census Bureau, "Calculating measures of error for derived estimates" in *ACS General Handbook* (U.S. Government Printing Office, Washington, DC, 2018).
75. J. Maples, "Calculating coefficient of variation for the minimum change school district poverty estimates and the assessment of the impact of nongecoded tax returns" (U.S. Census Bureau, Statistics 2008–10, 2008).
76. D. Campbell et al., "The robustness reproducibility of the American economic review" (I4R Discussion Paper Series, Working Paper 124, 2024).
77. T. Broderick, R. Giordano, R. Meager, An automatic finite-sample robustness metric: When can dropping a little data change conclusions? Part I: Definitions and experiments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **384**, 20250001 (2026).
78. A. Acquisti, "The economics of privacy at a crossroads" in *The Economics of Privacy*, A. Goldfarb, C. E. Tucker, Eds. (University of Chicago Press, Chicago, IL, 2024), pp. 21–72.
79. J. R. Cook, L. A. Stefanski, Simulation-extrapolation estimation in parametric measurement error models. *J. Am. Stat. Assoc.* **89**, 1314–1328 (1994).
80. D. H. Autor, D. Dorn, G. H. Hanson, The China syndrome: Local labor market effects of import competition in the United States. *Am. Econ. Rev.* **103**, 2121–2168 (2013).
81. C. Buntain, R. Bonneau, J. Nagler, J. A. Tucker, Measuring the ideology of audiences for web links and domains using differentially private engagement data. *Proc. Int. AAAI Conf. Web Social Media* **17**, 72–83 (2023).
82. C. Dwork et al., The reusable holdout: Preserving validity in adaptive data analysis. *Science* **349**, 636–638 (2015).
83. F. Echenique, K. He, Screening p-hackers: Dissemination noise as bait. *Proc. Natl. Acad. Sci.* **121**, e2400787121 (2024).
84. K. Martin, Privacy Myths and Mistakes: Paradoxes, tradeoffs, and the omnipotent consumer. SSRN [Preprint] (2024). <https://papers.ssrn.com/abstract=5022081> (Accessed 22 September 2026).
85. L. A. Barba, Terminologies for reproducible research. arXiv [Preprint] (2018). <http://arxiv.org/abs/1802.03311> (Accessed 20 November 2024).
86. K. Broman et al., "Recommendations to funding agencies for supporting reproducible research" (Tech. Rep., American Statistical Association, 2017).
87. A. Gelman, H. Stern, The difference between "Significant" and "Not Significant" is not itself statistically significant. *Am. Stat.* **60**, 328–331 (2006).
88. G. Cumming, Replication and p intervals: P values predict the future only vaguely, but confidence intervals do much better. *Perspect. Psychol. Sci. A: J. Assoc. Psychol. Sci.* **3**, 286–300 (2008).
89. C. Dwork, A. Roth, The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.* **9**, 211–407 (2013).
90. Y. Kang et al., Input perturbation: A new paradigm between central and local differential privacy. arXiv [Preprint] (2020). <https://arxiv.org/abs/2002.08570> (Accessed 16 October 2025).
91. D. Kifer, A. Machanavajjhala, "No free lunch in data privacy" in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data, SIGMOD '11* (Association for Computing Machinery, New York, NY, 2011), pp. 193–204.
92. P. Topalova, Factor immobility and regional impacts of trade liberalization: Evidence on poverty from India. *Am. Econ. J.: Appl. Econ.* **2**, 1–41 (2010).
93. A. D. Crane, A. Koch, L. Lin, Real effects of markets on politics: Evidence from US presidential elections. *Am. Econ. Rev. Insights* **6**, 73–88 (2024).
94. M. Weber, B. D. Baker, Does school funding matter in a pandemic? COVID-19 instructional models and school funding adequacy. *AERA Open* **11**, 23328584251327581 (2025).
95. C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution" in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* (University of California Press, 1956), pp. 197–207.
96. J. O. Berger, Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Stat.* **4**, 223–226 (1976).
97. R. Steed, E. A. S. Mustri, A. Acquisti, Code and Data from "Reproducibility of social science research using aggregate statistics with noise infused for differential privacy." GitHub. <https://github.com/ryansteed/dp-reproducibility>. Deposited 9 September 2026.