Statistically Valid Inferences from Privacy Protected Data*

Georgina Evans† Gary King‡
Margaret Schwenzfeier§ Abhradeep Thakurta¶

December 7, 2019

Abstract

Unprecedented quantities of data that could help social scientists understand and ameliorate the challenges of human society are presently locked away inside companies, governments, and other organizations, in part because of worries about privacy violations. We address this problem with a general-purpose data access and analysis system with mathematical guarantees of privacy for individuals who may be represented in the data, statistical guarantees for researchers seeking population-level insights from it, and protection for society from some fallacious scientific conclusions. We build on the standard of "differential privacy" but, unlike most such approaches, we also correct for the serious statistical biases induced by privacy-preserving procedures, provide a proper accounting for statistical uncertainty, and impose minimal constraints on the choice of data analytic methods and types of quantities estimated. Our algorithm is easy to implement, simple to use, and computationally efficient; we also offer open source software to illustrate all our methods.

*The current version of this paper is available at GaryKing.org/dp. Many thanks for helpful comments to Adam Breuer, Merce Crosas, Cynthia Dwork, Max Golperud, Andy Guess, Kosuke Imai, Dan Kifer, Solomon Messing, Xiao-Li Meng, Nate Persily, Aaron Roth, Adam Smith, Salil Vadhan, Sergey Yekhanin, and Xiang Zhou. Thanks also for help from the Alexander and Diviya Maguro Peer Pre-Review Program at Harvard's Institute for Quantitative Social Science.

†Ph.D. Candidate, Department of Government, Harvard University, 1737 Cambridge Street Cambridge, MA 02138; Georgina-Evans.com, GeorginaEvans@g.harvard.edu.

‡Albert J. Weatherhead III University Professor, Institute for Quantitative Social Science, 1737 Cambridge Street, Harvard University, Cambridge MA 02138; GaryKing.org, King@Harvard.edu.

§Ph.D. Candidate, Department of Government, Harvard University, 1737 Cambridge Street Cambridge, MA 02138; MegSchwenzfeier.com, schwenzfeier@g.harvard.edu.

¶Assistant Professor, Department of Computer Science, University of California Santa Cruz, bit.ly/AbhradeepThakurta, aguhatha@ucsc.edu.

1 Introduction

Just as more powerful telescopes empower astronomers, the accelerating influx of data about the political, social, and economic worlds has enabled considerable social science progress in understanding and ameliorating the challenges of human society. Yet, although we have more data than ever before, we may now have a smaller fraction of the data in the world than ever before because huge amounts are now locked up inside private companies, in part because of privacy concerns (King and Persily, In press). If we are to do our jobs as social scientists, we have no choice but to find ways of unlocking data from industry, as well as from governments, nonprofits, and other researchers. We might hope that government or society will take actions to support this mission, but we can also take responsibility ourselves and begin to develop technological solutions to these political problems.

In this paper, we develop methods to foster an emerging change in the paradigm for sharing research information. Under the familiar *data sharing regime*, trusted researchers simply obtain copies of data from others (perhaps with a data use agreement). Yet, with the public's increasing concerns over privacy, and data holders' (companies, governments, researchers, and others) desire to respond, this regime is failing. Fueling these concerns is the discovery that the common practice of de-identification does not reliably protect individual identities (Sweeney, 1997); nor does aggregation, query auditing, data clean rooms, legal agreements, restricted viewing, paired programmer models, and others (Dwork and Roth, 2014). And not only does the venerable practice of trusting researchers to follow the rules fail spectacularly at times (like the Cambridge Analytica scandal, sparked by a single researcher), but it turns out that even trusting a researcher who is known to be trustworthy does not always guarantee privacy (Dwork and Ullman, 2018).

An alternative approach that may help persuade some data holders to allow academic research is the *data access regime*. Under this regime, we begin with a trusted computer server that holds confidential data and treats researchers as potential "adversaries," meaning that they may try to learn individuals' private information while also seeking knowledge for research to generate public good. Then we add a "differentially private" algorithm that makes it possible for researchers to discover population-level insights but impossible to reliably detect the effect of the inclusion or exclusion of any one individual in the dataset or the value of any one person's variables. Under this data access regime researchers run statistical analyses on the server and receive "noisy" results computed by this privacy-preserving algorithm, but are limited by the total number of runs they may perform (so that they cannot repeat the same query many times and average away the noise). Differential privacy is a widely accepted mathematical standard for data access systems that promises to avoid some of the zero-sum policy debates over balancing the interests of individuals with the public good that can come from research. It also seems to satisfy regulators and others.[1]

A fast growing literature has formed around differential privacy, seeking to balance privacy and utility, but the current measures of "utility" provide little utility to social scientists or other statistical analysts. Statistical inference usually involves choosing a target *population* of interest, identifying the *data generation process*, and then using the resulting *dataset* to learn about features of the population. Valid inferences require methods with known statistical properties (such as unbiasedness, consistency, etc.) and honest assessments of uncertainty (e.g. standard errors). In contrast, privacy researchers typically begin with the choice of a target (confidential) *dataset*, add *privacy-protective procedures*, and then use the resulting *differentially private dataset or analyses* to infer to the confidential dataset — usually without regard to the data generation process or valid population inferences. This approach is useful for designing privacy algorithms but, as Wasserman (2012) puts it, "I don't know of a single statistician in the world who would analyze data this way."[2]

[1] Differential privacy was introduced by Dwork, McSherry, et al. (2006) and generalizes the social science technique of "randomized response" to elicit sensitive information in surveys (see Blair, Imai, and Zhou, 2015; Glynn, 2013; Warner, 1965); see Dwork and Roth (2014) and Vadhan (2017) for overviews and Wood et al. (2018) for a nontechnical introduction.

[2] "In statistical inference the sole source of randomness lies in the underlying model of data generation, whereas the estimators themselves are a deterministic function of the dataset. In contrast, differentially private estimators are inherently random *in their computation*. Statistical inference that considers *both* the randomness in the data and the randomness in the computation is highly uncommon" (Sheffet, 2017). As Karwa and Vadhan (2017) write, "Ignoring the noise introduced for privacy can result in wildly incorrect results at finite sample sizes. . . this can have severe consequences." On the essential role of inference and uncertainty and in science, see King, Keohane, and Verba (1994).

To make matters worse for social scientists, most privacy-protective procedures induce severe bias in population inferences, and even in inferences to features of confidential datasets. These include *adding random error*, which induces measurement error bias, and *censoring* (known as “clamping” in computer science), which induces selection bias (Blackwell, Honaker, and King, 2017; Stefanski, 2000; Winship and Mare, 1992). We have not found a single prior study that tries to correct for both (although some avoid the effects of censoring in theory at the cost of additional noise in practice; Karwa and Vadhan 2017; Smith 2011) and few have accurate uncertainty estimates. This is crucial because inferentially invalid data access systems can harm societies, organizations, and individuals — such as by inadvertently encouraging the distribution of misleading medical, policy, scientific, and other conclusions — even if it successfully protects individual privacy. For these reasons, using systems presently designed to ensure differential privacy would be unattractive for social science analysis.[3]

Social scientists and others need algorithms on data access systems with both inferential validity and differential privacy. We offer one such algorithm that is approximately unbiased, has lower variance than uncorrected estimates, and comes with accurate uncertainty estimates. The algorithm also turns censoring from a feature that severely biases statistical estimates in order to protect privacy to an attractive feature that greatly reduces the amount of noise needed to protect privacy while still leaving estimates approximately unbiased. The algorithm is easy to implement, computationally efficient even for very large datasets, and, because the entire dataset never needs to be stored in the same place, may offer additional security protections.

Our algorithm is *generic*, designed to minimally restrict the choice among statistical procedures, quantities of interest, data generating processes, and statistical modeling assumptions. Because the algorithm does not constrain researcher choices in these ways,

[3]Inferential issues also affect differential privacy applications outside of data access systems (the so called “global model”). These include “local model” systems where private calculations are made on a user’s system and sent back to a company in a way that prevents it from making reliable inferences about individuals — including Google’s Chrome (Erlingsson, Pihur, and Korolova, 2014) and their other core products (Wilson et al., 2019), Apple’s MacOS (Tang et al., 2017), and Microsoft’s Windows (Ding, Kulkarni, and Yekhanin, 2017) — and the US Census Bureau’s efforts to release differentially private datasets (Garfinkel, Abowd, and Powazek, 2018).

it may be especially well suited for building data access systems designed for research. When valid inferential methods exist or are developed for more restricted use cases, they may sometimes allow less noise for the same privacy guarantee. As such, one productive plan for building a general-purpose data access system may be to first implement our algorithm and to then gradually add these more specific approaches when they become available as preferred choices.[4]

We offer an introduction to differential privacy and describe the inferential challenges in analyzing data from a differentially private data access system, in Section 2. We give a generic differentially private algorithm in Section 3 which, like most such algorithms, is statistically biased. We therefore introduce bias corrections and variance estimators in Section 4 together with the private algorithm accomplishes our goals. We illustrate the performance of this approach in finite samples via Monte Carlo simulations in Section 5 and offer practical advice for implementation and use in Section 6. Appendices add technical details. We are also making available open source software (called UnbiasedPrivacy) to illustrate all the methods described herein.

2 Differential Privacy and its Inferential Challenges

We now define the differential privacy standard, describe its strengths, and highlight the challenges it poses for proper statistical inference. Throughout, we modify notation standard in computer science so that it is more familiar to social scientists.

2.1 Definitions

Begin with a confidential dataset D, defined as a collection of N rows of numerical measurements constructed so that each individual whose privacy is to be protected is repre-

[4]For example, Karwa and Vadhan (2017) develop finite sample confidence intervals with proper coverage for the mean of a normal density; Barrientos et al. (2019) offer differentially private significance tests for linear regression coefficients; Gaboardi et al. (2016) propose chi-squared tests for goodness of fit tests for multinomial data and independence between two categorical variables; Smith (2011) shows that, for a specific class of estimators and of data generating processes, there exists a differentially private estimator with the same asymptotic distribution; Wang, Lee, and Kifer (2015) propose accurate p-values for chi-squared tests of independence between two variables in tabular data; Wang, Kifer, and Lee (2018) develop differentially private confidence intervals for objective or output perturbation; and Williams and McSherry (2010) provide an elegant marginal likelihood approach for moderate sized datasets.

sented in at most one row.[5]

Statistical analysts would normally calculate a statistic s (such as a count, mean, parameter estimate, etc.) from D as a fixed number, say $s(D)$. For inference, they then conceptualize $s(D)$ as a random variable given (hypothetical, unobserved) repeated draws of D following the same data generation process from a population. In contrast, privacy analysts ignore populations and data generation processes and treat $s(D)$ as the fixed unobserved quantity of interest. They then construct a "mechanism" $M(s, D)$, which is a privacy-protected version of the same statistic, calculated by injecting carefully calibrated noise and censoring at some point before returning the result. As we will show, the specific types of noise and censoring are specially designed to satisfy differential privacy. Privacy researchers conceptualize (hypothetical, unobserved) sampling distributions of $M(s, D)$, but these are generated by repeated draws of the noise from the same distribution, with D fixed.

Meeting the differential privacy standard prevents a researcher from reliably learning anything different from a dataset regardless of whether an individual has been included or excluded. To formalize this notion, consider two datasets D and D' that differ in at most one row (a maximum Hamming distance of 1). Then, the standard requires that the probability (or probability density) of any analysis result m from dataset D, $\Pr[M(s, D) = m]$, be *indistinguishable* from the probability that the same result is produced by the same analysis of dataset D', $\Pr[M(s, D') = m]$, where the probabilities take D as fixed and are computed over the noise.

We write an intuitive version of the differential privacy standard (using the fact that $e^\epsilon \approx 1 + \epsilon$ for small ϵ) by defining "indistinguishable" as the ratio of the probabilities falling within ϵ of equality (which is 1). Thus, a mechanism is said to be ϵ-differentially private if

$$\frac{\Pr[M(s, D) = m]}{\Pr[M(s, D') = m]} \in 1 \pm \epsilon, \tag{1}$$

where ϵ is a pre-chosen level of possible privacy leakage, with smaller values potentially

[5] Hierarchical data structures, or dependence among units, is allowed within but not between rows. For example, rows could represent a family with variables for different family members.

giving away less privacy (by requiring more noise or censoring).[6] Many variations and extensions of Equation 1 have been proposed (Desfontaines and Pejó, 2019). We use the most popular, known as "(ϵ, δ)-differential privacy" or "approximate differential privacy," which adds a very small chosen offset δ to the numerator of the ratio in Equation 1. This second privacy parameter, which the user chooses such that $\delta < 1/N$, turns out to allow mechanisms with (statistically convenient) Gaussian noise processes. This relaxation also has Bayesian interpretations, with the posterior distribution of $M(s, D)$ close to that of $M(s, D')$, and also that an (ϵ, δ)-differentially private mechanism is ϵ-differentially private with probability at least $1 - \delta$ (Vadhan, 2017, p.355ff). We can also express approximate differential privacy more formally as requiring that each of the probabilities be bounded by a linear function of the other:[7]

$$\Pr[M(s, D) = m] \quad \leq \quad \delta + e^{\epsilon} \cdot \Pr[M(s, D') = m]. \tag{2}$$

Consistent with political science research showing that secrecy is best thought of on a continuum (Roberts, 2018), the differential privacy standard quantifies the privacy leakage of a given mechanism via the choices of ϵ and δ. Differential privacy is expressed in terms of the maximum possible privacy loss, but the expected privacy loss is considerably less than this worst case analysis, often by orders of magnitude (Carlini et al., 2019; Jayaraman and Evans, 2019). It also protects small groups in the same way as individuals, with the maximum risk $k\epsilon$ dropping linearly in group size k. And because mechanisms with different small values of ϵ have similar properties, even some violations of the differential privacy standard may still be differentially private for larger values of ϵ and δ.

[6]Using this multiplicative (ratio) metric to indicate what is "indistinguishable" turns out to be much more protective of individual privacy than some others, such as an additive (difference) metric. For example, consider an obviously unacceptable mechanism: "choose one individual uniformly at random and disclose all of his or her data." This mechanism is not differentially private (the ratio can be infinite and thus greater than any finite ϵ), but it may seem safe on an additive metric because the impact of adding or removing one individual on the difference in the probability distribution of a statistical output is proportional to at most $1/N$.

[7]Our algorithms below also satisfy a strong version of approximate differential privacy known as Rényi differential privacy; see Mironov (2017).

2.2 Example

The literature includes many differentially private mechanisms. These add noise and censoring to the data inputs, the output estimates, or various parts of internal calculations, such as the gradients, elements of $X'X$ matrices for regression, or others. For the goal of developing a generic algorithm, we now introduce the differentially private *Gaussian mechanism*, which we use throughout this paper. This mechanism, like most others, is statistically biased and inconsistent and does not come with uncertainty estimates, but we will use a version of it in Section 3 to build a differentially private algorithm and then provide corrections in Section 4 to make it statistically valid.

To fix ideas, consider a confidential database of the incomes of people in a neighborhood of Seattle and the mean income as the target quantity of interest: $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$. Given the researcher's choice of privacy parameters ϵ and δ and bounding parameter $\Lambda > 0$, we define this as the censored mean plus Gaussian noise:

$$M(\text{mean}, D) = \hat{\theta} + \mathcal{N}(0, S^2) \tag{3}$$

with mean $\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} c(y_i, \Lambda)$ and censoring function

$$c(y, \Lambda) = \begin{cases} y & \text{if } y \in [-\Lambda, \Lambda] \\ \text{sgn}(y)\Lambda & \text{if } y \notin [-\Lambda, \Lambda]. \end{cases} \tag{4}$$

The remaining question is how much noise to add or, in other words, the definition of $S \equiv S(\Lambda, \epsilon, \delta, N)$. Under approximate differential privacy, we add only as much noise as necessary to satisfy Equation 2, which we write for this purpose as $\mathcal{N}(t \mid \hat{\theta}, S^2) \leq \delta + e^{\epsilon} \cdot \mathcal{N}(t \mid \hat{\theta} + \Delta, S^2)$, where Δ is the sensitivity of this estimator (the largest change over all possible pairs of datasets that differ by at most one row), where $|\hat{\theta}_D - \hat{\theta}_{D'}| \leq \Delta$ such that $\hat{\theta}_D$ and $\hat{\theta}_{D'}$ denote the estimator computed from D and D', respectively. The censored mean $\hat{\theta}$ has sensitivity $\Delta = 2\Lambda/N$.

Although in practice, we recommend a more general solution with a tight bound,[8] the

[8]Dwork and Roth (2014) show that Equation 5 holds only for $\epsilon \leq 1$. For any value of ϵ, we recommend the tight bound based on the numerical solution by Balle and Wang (2018) which allows much smaller values of S even when $\epsilon \leq 1$. As a brief summary, note that if M is Gaussian, then Equation 2 can be written in terms of the cumulative standard normal density: $\Phi\left(\frac{\Delta}{2\sigma} - \frac{\epsilon\sigma}{\Delta}\right) - e^{\epsilon}\Phi\left(-\frac{\Delta}{2\sigma} - \frac{\epsilon\sigma}{\Delta}\right) \leq \delta$. Then set S equal to the minimum σ satisfying this inequality. A clear benefit is that the solution exactly calibrates the noise to a given privacy budget and hence minimizes the variance of the output for a given level of $\{\epsilon, \delta\}$. The cost of this approach is computing a unique numerical solution for each run of their algorithm.

simplest solution gives

$$S \equiv S(\Lambda, \epsilon, \delta, N) = \frac{\Delta\sqrt{2\ln(1.25/\delta)}}{\epsilon} = \frac{2\Lambda\sqrt{2\ln(1.25/\delta)}}{N\epsilon}. \tag{5}$$

For intuition, we also simplify further with an arbitrary but convenient choice for δ:

$$S(\Lambda, \epsilon, 0.0005, N) \approx \frac{8\Lambda}{N\epsilon}. \tag{6}$$

Equation 6 shows that, to protect the biggest possible outlier, differential privacy allows us to add less noise if each person is submerged in a sea of many others (larger N), if less privacy is required (larger ϵ), or if more censoring is used (smaller Λ).

With any level of censoring, $\hat{\theta}$ is obviously a biased estimate of $\bar{y}$: $E(\hat{\theta}) \neq \bar{y}$. To reduce censoring and thus bias, we can choose larger values of Λ but that unfortunately would increase the noise, the statistical variance, and our uncertainty estimates; similarly, reducing noise by choosing a smaller value of Λ increases the impact of censoring. These choices are important: if the largest outlier could be Bill Gates or Jeff Bezos, then avoiding censoring would require adding so much noise that it would swamp any relevant signal from the data. We resolve much of this tension Section 4.

2.3 Inferential Challenges

We now discuss four issues the tools of differential privacy pose from the perspective of statistical inference. For some we offer corrections; for others, we suggest how to adjust statistical analysis practices.

First, censoring data induces selection bias in statistical inference. Avoiding censoring by adding more noise is no solution because any amount of noise induces bias in statistical estimators for nonlinear functions of the data. Moreover, even for estimators that are unbiased (like the mean in Section 2.2), the added noise makes unadjusted standard errors statistically inconsistent. Ignoring either measurement error bias or selection bias is usually a major inferential mistake and may change substantive conclusions, the properties of estimators, and the validity of uncertainty estimates, often in negative, unknown, or surprising ways.[9]

[9]For example, adding mean-zero noise to one variable or its mean induces no bias in estimating the

Second, uncertainty estimators are rarely proposed or even discussed in the literature on differentially private mechanisms. Unfortunately, accurate uncertainty estimates cannot be generated by using differentially private versions of classical uncertainty estimates. To be more specific, in a system without differential privacy, let $\hat{\theta}$ be a point estimate in an observed dataset of a quantity of interest θ from an unobserved population. Denote by $V(\hat{\theta})$ the variance of $\hat{\theta}$ over repeated (hypothetical, unobserved) samples of datasets drawn with the same data generation process from the population, with an estimate from the one observed dataset $\hat{V}(\hat{\theta})$, its square root being the standard error, a commonly reported measure of statistical uncertainty. For a proper scientific statement, it would be sufficient to have (1) an estimator $\hat{\theta}$ with known statistical properties (such as unbiasedness, consistency, efficiency, etc.), and (2) a variance estimate $\hat{V}(\hat{\theta})$ that reflects the true variance $V(\hat{\theta})$. Consider now the differentially private point estimate $\hat{\theta}^{\text{dp}}$. Although it would be easy to compute a differentially private variance estimate $\hat{V}(\hat{\theta})^{\text{dp}}$ using the same type of mechanism, it is of no direct use, since $\hat{\theta}$ is never disclosed and so its variance is irrelevant. Indeed, $\hat{V}(\hat{\theta})^{\text{dp}}$ is a biased estimate of the relevant uncertainty, $V(\hat{\theta}^{\text{dp}})$. Since we will show below that $\hat{\theta}^{\text{dp}}$ itself is biased, $V(\hat{\theta}^{\text{dp}})$ would be of no direct use even if it were known.

Third, to avoid researchers rerunning the same analysis many times and averaging away the noise, their analyses must be limited. This limitation is formalized via a differential privacy property known as *composition*: If mechanism k is (ϵ_k, δ_k)-differentially private, for $k = 1, \ldots, K$, then disclosing all K estimates is $(\sum_{k=1}^{K} \epsilon_k, \sum_{k=1}^{K} \delta_k)$-differentially private.[10] Then the restriction is implemented via a quantitative *privacy budget* in which the data provider allocates a *total* value of ϵ to a researcher who can then divide it up and run as many analyses, of whatever type, as they choose, so long as the sum of all the ϵs

population mean, but adding noise to its variance creates a biased estimate. The estimated slope coefficient in a simple regression of y on x where, with random measurement error in x, is biased toward zero; if we add variables with or without measurement error to this regression, the same coefficient can be biased by any amount and in any direction. Censoring sometimes attenuates causal effects but it can also exaggerate them; predictions and estimates of other quantities of interest can be too high, too low, or have their signs changed. Measurement error and censoring should not be ignored; they do not come out in the wash.

[10]Alternatively, if the K quantities are disclosed simultaneously and returned in a batch, then we could choose to set the variance of the error for all together at a higher individual level but lower collective level, dependent also on the type of noise added (Bun and Steinke, 2016; Mironov, 2017).

across all their analyses does not exceed their total privacy budget.

This strategy has the great advantage of enabling each researcher to make these choices, rather than a central authority such as the data provider. However, when the total privacy budget is used up, no researcher can ever run a new analysis on the same data unless the data provider chooses to increase the budget. This constraint is a useful feature to protect privacy, but it utterly changes the nature of statistical analysis. To see this, note that best practice recommendations have long included trying to avoid being fooled *by the data* — by running every possible diagnostic, fully exploring the dataset, and conducting numerous statistical checks — and *by the researcher's personal biases* — such as by preregistration ex ante to constrain the number of analyses that will be run to eliminate "p-hacking" or correcting for "multiple comparisons" ex post (Monogan, 2015). One obviously needs to avoid being fooled in any way, and so researchers normally try to balance the resulting contradictory advice to avoid each problem. In contrast, differential privacy tips the scales: Remarkably, it makes solving the second problem almost automatic (Dwork, Feldman, et al., 2015), but it also severely reduces the probability of serendipitous discovery and increases the odds of being fooled by unanticipated data problems. Successful data analysis with differential privacy thus requires careful planning, although less stringently than with pre-registration.

Finally, a researcher can learn about an individual from a differentially private mechanism, but no more than if that individual were excluded from the data set. For example, suppose research indicates that women are more likely to share fake news with friends on social media than men; then, if you are a woman, everyone knows that you have a higher risk of sharing fake news. But the researcher would have learned this population-level fact whether or not you were included in the dataset and so you have no reason to withhold your information.

However, we must also be certain that the differentially private mechanism is inferentially valid. If researchers use privacy preserving mechanisms that bias statistical procedures and no corrections are applied, society can sometimes be mislead and numerous individuals can be hurt by publishing incorrect population level inferences. (In fact, it is older people, not women, who are more likely to share fake news! See Guess, Nagler, and Tucker 2019.) Fortunately, all of differential privacy's properties are preserved under post-processing, which means that no privacy loss will occur when, below, we correct for inferential biases induced by noise and censoring (or if results are published or mixed with any other data sources). In particular, for any data analytic function f not involving private data D, if $M(s, D)$ is differentially private, then $f[M(s, D)]$ is differentially private, regardless of assumptions about potential adversaries or threat models.

Although differential privacy may seem to follow a "do no *more* harm" principle, careless use of this technology can in fact harm individuals and society if we do not also ensure inferential validity. The biases from ignoring measurement error and selection can each separately or together reverse, attenuate, exaggerate, or nullify statistical results. Helpful medical treatments could be discarded. Harmful practices may be promoted. Of course, when providing access to confidential data, not using differential privacy may also have grave costs to individuals. Data providers must therefore ensure that data access systems are *both* differentially private and inferentially valid.

3 A Generic Differentially Private Estimator

Our approach, which like our software we call UnbiasedPrivacy (or UP), has two parts — a differentially private mechanism introduced in this section and a bias correction of the differentially private result, using the algorithm in Section 4. Section 4 also gives accurate uncertainty estimates in the form of standard errors.

Let $\mathbf{D}$ denote a *population* data matrix, from which N observations are selected to form our observed data matrix D. Our goal is to estimate some (fixed scalar) quantity of interest, $\theta = s(\mathbf{D})$ with the researcher's choice of statistical procedure s (among the vast array that are statistically valid under bootstrapping, i.e., any statistic with a positive bounded second Gateaux derivative and Hadamard differentiability; see Wasserman 2006, p.35). Let $\hat{\theta} = s(D)$ denote an estimate of θ computed from the private data in the way we normally would without privacy protective procedures. Because privacy concerns prevent $\hat{\theta}$ from being disclosed, we show here instead how to estimate a differentially private estimate of θ denoted $\hat{\theta}^{\text{dp}}$ which, like many such estimators, is substantially biased but is bias corrected in the next section.

To estimate, $\hat{\theta}^{\text{dp}}$, the user chooses a statistical method (logit, regression, cross-tabulation, etc.), a quantity of interest estimated from the statistical method (causal effect, risk difference, predicted value, etc.), and values for each of the privacy parameters, Λ, ϵ, and δ (see Section 6 for advice on making these choices).

We give the details of our proposed mechanism $M(s, D) = \hat{\theta}^{\text{dp}}$ in Section 3.1. It uses a partitioning version of the "sample and aggregate" algorithm (Nissim, Raskhodnikova, and Smith, 2007), to ensure differential privacy for almost any statistical method and quantity of interest. We also incorporate an optional application of the computationally efficient "bag of little bootstraps" algorithm (Kleiner et al., 2014), that will ensure an aspect of inferential validity generically, by not having to worry about differences in how to scale up different statistics from each partition to the entire dataset.

3.1 Mechanism

Randomly partition rows of D as $\{D_1, \ldots, D_P\}$, each of subset size $n \approx N/P$ (we discuss the choice of P below), and then follow this algorithm.

1. For partition p ($p = 1, \ldots, P$)
   (a) Compute an estimate $\hat{\theta}_p$ (of a quantity of interest θ) *either* directly (being careful to appropriately scale up quantities that require it; see Section 3.3) *or* by bootstrapping (where scaling up is automatic) as follows:
      i. For bootstrap simulation b ($b = 1, \ldots, B$)
         A. Simulate bootstrap b by sampling one weight for each of the n units in partition p as: $w_b \equiv \{w_{1,b}, \ldots, w_{n,b}\} \sim \text{Multinomial}(N, \mathbf{1}_n/n)$.
         B. Calculate a statistic (an estimate of population value θ) from bootstrapped sample b in partition p: $\hat{\theta}_{p,b} = s(D_p, w_b)$, such as a predicted value, expected value, or classification.
      ii. Summarize the set of bootstrapped estimates within each partition with an (unobserved, nonprivate) *estimator*, which we write generically as $\hat{\theta}_p$.

Examples include: a *point estimator*, such as the mean $\hat{\theta}_p = m(\hat{\theta}_{p,b})$ or the probability of the Democrat winning a majority of the vote, $\hat{\theta}_p = m[\mathbb{1}(\hat{\theta}_{p,b} > 0.5)]$, or an *uncertainty estimator*, such as the variance $\hat{\theta}_p = v(\hat{\theta}_{p,b})$ (the squared standard error).

(b) For a fixed value of of the bounding parameter $\Lambda > 0$ chosen ex ante, *censor* the estimate $\hat{\theta}_p$ as $c(\hat{\theta}_p, \Lambda)$ using Equation 4.

2. Form a differentially private estimate $\hat{\theta}^{\text{dp}}$ using a version of the Gaussian mechanism (Section 2.2): average the nonprivate estimates (over partitions) and add appropriately calibrated noise:

$$\hat{\theta}^{\text{dp}} = \hat{\theta} + e \tag{7}$$

where

$$\hat{\theta} = \frac{1}{P}\sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda), \qquad e \sim \mathcal{N}(0, S_{\hat{\theta}}^2), \qquad S_{\hat{\theta}} = S(\Lambda, \epsilon, \delta, P), \tag{8}$$

and S is defined in Equation 5.

3.2 Privacy Properties

Privacy is ensured in this algorithm by each individual appearing in at most one partition, and by the censoring and noise in the aggregation mechanism ensuring that data from any one individual can have no measurable effect on the distribution of possible outputs. Each partition can even be sequestered on a separate server, which may reduce security risks.

The advantage of always using the mean over partitions, rather than another aggregation procedure, in the expression for $\hat{\theta}$ is that $S_{\hat{\theta}}$ can be calibrated generically to the sensitivity of this (censored) mean rather than having to derive the sensitivity anew for each estimator. The cost of this strategy is additional noise because P rather than N appears in the denominator of the variance. Thus, from the perspective of reducing noise, P should be set as large as possible, subject to the constraints that (1) the number of units in each partition $n \approx N/P$ gives valid statistical results in each bootstrap and the estimate being sensible (such as regression covariates being of full rank) and (2) n is large enough and growing faster than P (to ensure the central limit theorem can be applied). (Subsampling itself increases the variance of nonlinear estimators, but usually much less than the increased variance due to privacy protective procedures; see also Mohan et al. (2012) for more formal methods of optimizing P.)

3.3 Inferential Properties

The statistical properties of estimators from our algorithm differ depending on type. For example, consider two conditions: (1) an assumption we maintain until the next section that Λ is large enough so that censoring has no effect ($c(\hat{\theta}_p, \Lambda) = \hat{\theta}_p$), and (2) an estimator applied to the private data that is unbiased for the chosen quantity of interest. If these two conditions hold, and because also we are injecting additive noise at the last stage of the mechanism (rather in the data, the objective function, sequential steps, etc.), the point estimates are unbiased:

$$E(\hat{\theta}^{\text{dp}}) = \frac{1}{P}\sum_{p=1}^{P} E(\hat{\theta}_p) + E(e) = \theta. \tag{9}$$

In practice, however, choosing the bounding parameter Λ involves a bias-variance trade-off: If Λ is set to the maximum possible sensitivity, censoring has no effect and $\hat{\theta}^{\text{dp}}$ is unbiased, but the noise is large (see Equation 8). Choosing smaller values of Λ reduce noise, which reduces the variance of the estimator, but it simultaneously increases bias due to censoring (Section 3, Step 1b). Also, if the estimator is applied without privacy protective procedures is biased (such as by violating statistical assumptions or merely being a nonlinear function of the data), then our algorithm will not magically remove the bias, but it will not add bias.

In contrast, uncertainty estimators require adjustment even if the three conditions are met. For example, the variance of the differentially private estimator is $V(\hat{\theta}^{\text{dp}}) = V(\hat{\theta}) + S^2_{\hat{\theta}}$, but its naive variance estimator (the differentially private version of an unbiased nonprivate variance estimator) is biased:

$$E\left[\hat{V}(\hat{\theta})^{\text{dp}}\right] = E\left[\hat{V}(\hat{\theta}) + e\right] = V(\hat{\theta}) + E(e) = V(\hat{\theta}) \neq V(\hat{\theta}^{\text{dp}}). \tag{10}$$

Fortunately, we can compute an unbiased estimate of the variance of the differentially private estimator by simply adding back in the (known) variance of the noise: $\hat{V}(\hat{\theta}^{\text{dp}}) =$

$\hat{V}(\hat{\theta})+S^2_{\hat{\theta}}$, which is unbiased: $E\left[\hat{V}(\hat{\theta}^{\text{dp}})\right] = V(\hat{\theta}^{\text{dp}})$. Of course, because (2) will typically be violated, and the resulting censoring will bias our estimates, we must bias correct this estimate and then compute the variance of the corrected estimate, which will ordinarily require a more complicated expression.

Finally, as indicated in the algorithm, the "little bag of bootstraps" in algorithm step 1(a) can be replaced with the choice of an unbiased estimator applied directly to data within partition p, if estimators are scaled up appropriately by modifying Equation 7 to match the size of the entire dataset (see Politis, Romano, and Wolf, 1999). On the one hand, the bootstrap approach is simpler in two situations. First, the necessary scale factors the researcher would need to derive without bootstrapping differ depending on the type of estimator and quantity of interest; some quantities, like the variance, are a linear function of N, but each partition's estimate is a function of n and so need to be scaled up by a factor of $(N-n)/n$. In contrast, the bootstrapping approach can be applied generically, the same for all. Second, bootstrapping allows simpler and more flexible estimation strategies for some quantities, such as the probability that a causal effect is greater than zero, estimated by merely counting the proportion of positive bootstrap estimates that meet selected criteria. (It works well for estimating α_2 in the model below when P is unavoidably small and so would otherwise create grouping error.) On the other hand, some quantities that without differential privacy would be easy to compute with bootstrapping are difficult or impossible with differential privacy, such as the variance of post-processed quantities, like our bias corrected estimator.[11]

4 Valid Statistical Inference from Private Data

We develop here an approach to valid inference from private data using a post-processed version of the generic differentially private estimator developed in Section 3, which means it retains all of its privacy preserving properties. The post-processing bias corrects $\hat{\theta}^{\text{dp}}$ for

[11] We also note that setting $B = 1$ (for a single bootstrap) would save computation but would make some quantities (such as the standard error or a predicted probability of a plurality vote for a candidate) impossible to estimate by bootstrapping and would be inefficient even for those that are possible; alternatively, replacing the weights in Step 1(a)i,A with a large dataset of size N constructed from each partition would satisfy the scaling issue but would require P times as much memory for each of the P computations.

censoring, and results in our estimator, $\tilde{\theta}^{\text{dp}}$. We know of no prior attempt to correct for biases due to censoring in differentially private mechanisms. This bias correction has the effect of reducing the impact of whatever value of Λ is selected, and allows users to choose smaller values to reduce variance and use less of the privacy budget without bias concerns (see also Section 6). It even turns out that the variance of this bias corrected estimate is actually smaller than the uncorrected estimate, which is unusual for bias corrections. We also offer an estimate of this variance, denoted $\hat{V}(\tilde{\theta}^{\text{dp}})$.

4.1 Bias Correction

Our goal here is to correct the bias due to censoring in our estimate of θ. Figure 1 helps visualize the underlying distributions and notation we will introduce, with the original uncensored distribution in blue and the censored distribution in orange, which is made up of an unnormalized truncated distribution and the spikes (which replace the area in the tails) at $-\Lambda$ and Λ.

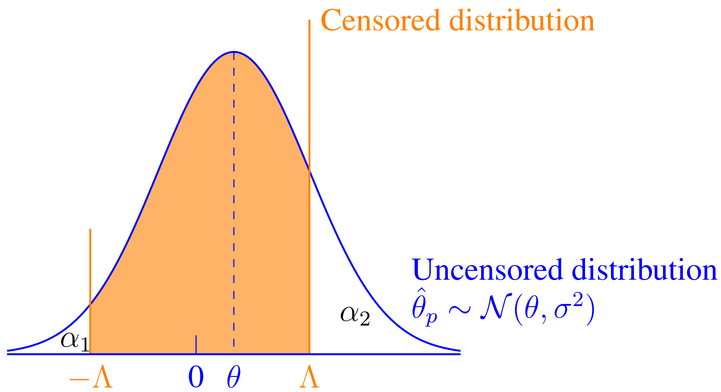

Figure 1: Underlying Distributions, before estimation. The censored distribution includes the orange area and the spikes at $-\Lambda$ and Λ.

Begin by assuming that n is large enough for the central limit theorem to apply (so that n grows faster than P; in practice $n > 30$ is usually sufficient), which means that the distribution of $\hat{\theta}_p$ across partitions, before censoring at $[-\Lambda, \Lambda]$, is $\mathcal{N}(\theta, \sigma^2)$, with the proportion left and right censored, respectively,

$$\alpha_1 = \int_{-\infty}^{-\Lambda} \mathcal{N}(t \mid \theta, \sigma^2) dt, \qquad \alpha_2 = \int_{\Lambda}^{\infty} \mathcal{N}(t \mid \theta, \sigma^2) dt. \tag{11}$$

We then write the expected value of $\hat{\theta}^{\text{dp}}$ as the weighted average of the mean of the truncated normal and of the spikes at $-\Lambda$ and Λ:

$$E\left(\hat{\theta}^{\text{dp}}\right) = -\alpha_1\Lambda + (1-\alpha_2-\alpha_1)\theta_T + \alpha_2\Lambda, \tag{12}$$

with truncated normal mean

$$\theta_T = \theta + \frac{\frac{\sigma}{\sqrt{2\pi}}\left[\exp\left(-\frac{1}{2}\left(\frac{-\Lambda-\theta}{\sigma}\right)^2\right) - \exp\left(-\frac{1}{2}\left(\frac{\Lambda-\theta}{\sigma}\right)^2\right)\right]}{1-\alpha_2-\alpha_1}. \tag{13}$$

After substituting our point estimate $\hat{\theta}^{\text{dp}}$ for the expected value in Equation 12, we are left with three equations (12 and the two in 11) and four unknowns (θ, σ, α_1, and α_2). We therefore use some of the privacy budget to obtain an estimate of α_2 (or to increase numerical stability we can try to estimate the larger of α_1 or α_2). Then we have three equations and three unknowns (θ, σ, α_1), which our open source software solves with a fast numerical solution, giving $\tilde{\theta}^{\text{dp}}$, which is our goal, the approximately unbiased estimate of θ, along with estimates $\hat{\sigma}_{\text{dp}}$ and $\hat{\alpha}_1^{\text{dp}}$.

For estimation, we note that α_2 is bounded to the unit interval, and so we could set $\Lambda_\alpha = 1$ without risk of censoring, but this would overstate this parameter's sensitivity by a factor of two. Instead of resolving this issue by changing the expression for censoring and S, we do so more conveniently, without changing the notation (or code) above, by simply reparameterizing as $\beta = \alpha_2 - 0.5$, setting $\Lambda_\beta = 0.5$, estimating and disclosing β, and then solving to obtain our estimate of α_2 before using it to solve our three equations.

4.2 Variance Estimation

We now derive a procedure for computing an estimate of the variance of our estimator, $\hat{V}(\tilde{\theta}^{\text{dp}})$, without any additional privacy budget expenditure. We have the two directly estimated quantities, $\hat{\theta}^{\text{dp}}$ and $\hat{\alpha}_2^{\text{dp}}$, and the three (deterministically post-processed) functions of these computed during bias correction: $\tilde{\theta}^{\text{dp}}$, $\hat{\sigma}^2_{\text{dp}}$, and $\hat{\alpha}_1^{\text{dp}}$. We then use standard simulation methods (King, Tomz, and Wittenberg, 2000): We treat the estimated quantities as random variables, bias correct to generate the others, and take the sample variance of the simulations of $\tilde{\theta}^{\text{dp}}$.

Thus, to represent estimation uncertainty, we draw the random quantities from a multivariate normal with plug-in parameter estimates. Using notation (i) to denote the ith simulation, we write:

$$\hat{\theta}^{\mathrm{dp}}(i), \hat{\alpha}_2^{\mathrm{dp}}(i) \sim \mathcal{N}\left(\begin{bmatrix}\hat{\theta}^{\mathrm{dp}} \\ \hat{\alpha}_2^{\mathrm{dp}}\end{bmatrix}, \begin{bmatrix}\hat{V}(\hat{\theta}^{\mathrm{dp}}) & \widehat{\mathrm{Cov}}(\hat{\alpha}_2^{\mathrm{dp}}, \hat{\theta}^{\mathrm{dp}}) \\ \widehat{\mathrm{Cov}}(\hat{\alpha}_2^{\mathrm{dp}}, \hat{\theta}^{\mathrm{dp}}) & \hat{V}(\hat{\alpha}_2^{\mathrm{dp}})\end{bmatrix}\right). \tag{14}$$

To implement this procedure we require $\hat{V}(\hat{\theta}^{\mathrm{dp}})$, $\hat{V}(\hat{\alpha}_2^{\mathrm{dp}})$, and $\widehat{\mathrm{Cov}}(\hat{\alpha}_2^{\mathrm{dp}}, \hat{\theta}^{\mathrm{dp}})$, which we show in Appendix A can be written as functions of information already disclosed. We plug these into Equation 14 and repeatedly draw $\{\hat{\theta}^{\mathrm{dp}}(i), \hat{\alpha}_2^{\mathrm{dp}}(i)\}$, each time bias correcting via the procedure in Section 4.1 to compute $\tilde{\theta}^{\mathrm{dp}}(i)$. Finally, we compute the sample variance over these simulations to yield our estimate $\hat{V}(\tilde{\theta}^{\mathrm{dp}})$.

5 Simulations

In this section, we evaluate the finite sample properties of our estimator via simple Monte Carlo simulations. We show that while (uncorrected) differentially private point estimates are inferentially invalid, our (bias corrected) estimators are approximately unbiased (when the non-private estimator is unbiased), and come with accurate uncertainty estimates. In addition, in part because our bias correction uses an additional disclosed parameter estimate ($\hat{\alpha}_2$), the variance of our estimator is usually lower than the variance of the uncorrected estimator.

The results appear in Figure 2, which we discuss after first detailing the data generation process. For four different types of simulations (in separate panels of the figure), we draw data for each row i from an independent linear regression model: $y_i \sim \mathcal{N}(1 + 3x_i, 10^2)$, with $x_i \sim \mathcal{N}(0, 7^2)$ drawn once and fixed across simulations. Our chosen quantity of interest is the coefficient on x_i with value $\theta = 3$. We study the bias of the (uncorrected) differentially private estimator $\hat{\theta}^{\mathrm{dp}}$, and our corrected version, $\tilde{\theta}^{\mathrm{dp}}$, as well as their standard errors.[12]

[12] We have tried different parameter values, functional forms, and distributions, all of which led to the same substantive conclusions. In Figure 2, for censoring (top left panel), we let $\alpha_1 = 0$, $\alpha_2 = \{0.1, 0.25, 0.375, 0.5, 0.625, 0.75\}$, $N = 100,000$, $P = 1,000$, and $\epsilon = 1$. For privacy (in the top right panel) and standard errors (bottom right), let $\epsilon = \{0.1, 0.15, 0.20, 0.30, 0.50, 1\}$, while setting $N = 100,000$, $P = 1,000$, and with Λ set so that $\alpha_1 = 0$ and $\alpha_2 = 0.25$. For sample size (bottom

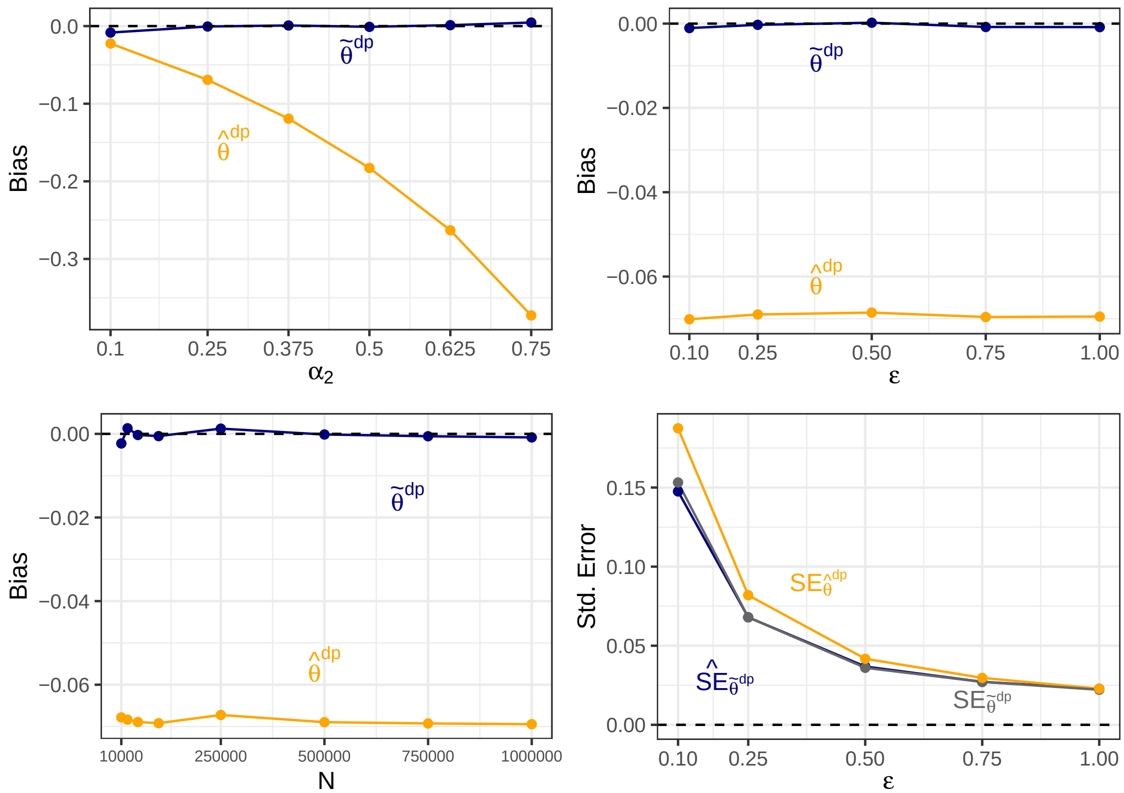

Figure 2: Monte Carlo Simulations

Begin with the top left panel, which plots bias on the vertical axis (with zero bias indicated by a dashed horizontal line at zero near the top) and the degree of censoring on the horizontal axis increasing from left to right (quantified by α_2). The orange line in this panel vividly shows how statistical bias in the (uncorrected) differentially private estimator $\hat{\theta}^{\text{dp}}$ sharply increases with censoring. In contrast, our (bias corrected) estimate in blue $\tilde{\theta}^{\text{dp}}$ is approximately unbiased regardless of the level of censoring.

The top right and bottom left panels also plot bias on the vertical axis with zero bias indicated by a horizontal dashed line. The bottom left panel shows the bias in the uncorrected estimate (in orange) for sample sizes from 10,000 to 1 million, and the top right panel shows the same for different values of ϵ. Our corrected estimate (in blue) is approximately zero in both panels, regardless of the value of N or ϵ.

Finally, the bottom right panel reveals that the standard error of $\tilde{\theta}^{\text{dp}}$ is approximately

left), set $N = \{10,000, 25,000, 50,000, 100,000, 250,000, 50,000, 100,000\}$, $P = 1,000$, $\epsilon = 1$, and determine Λ so that $\alpha_1 = 0$ and $\alpha_2 = 0.25$. We ran 1,000 Monte Carlo simulations except for $N \leq 50,000$ where we ran 4,000, and 2,000 for $\epsilon = 0.25$.

correct (i.e., equal to the true standard deviation across estimates, which can be seen because the blue and gray lines are almost on top of one another). It is even smaller for most of the range than the standard error of the uncorrected estimate $\hat{\theta}^{\text{dp}}$.

These simulations suggest that $\tilde{\theta}^{\text{dp}}$ is to be preferred to $\hat{\theta}^{\text{dp}}$ with respect to bias and variance in finite samples.

6 Practical Suggestions

Like any data analytic approach, how the methods proposed here are used in practice can be as important as their formal properties. We thus offer some practical suggestions for users and software designers. We discuss issues of reducing the societal risks of differential privacy, choosing ϵ, choosing Λ, theory and practice differences, and software design.

Reducing Differential Privacy's Societal Risks Data access systems with differential privacy are designed to reduce privacy risks to individuals. Correcting the biases due to noise and censoring, and adding proper uncertainty estimates, greatly reduces the remaining risks of the procedure to researchers and, through their results, to society. There is, however, another risk we must tackle: Consider a firm seeking public relations benefits by making data available for academics to create public good but, concerned about bad news for the firm that might come from the research, takes an excessively conservative position on the total privacy budget. In this situation, the firm would effectively be providing a big pile of useless random numbers while claiming public credit for making data available. No public good could be created, no bad news for the firm could come from the research results because all causal estimates would be approximately zero, and still the firm would benefit from great publicity.

To avoid this unacceptable situation, we now show how to estimate the statistical cost of differential privacy so we can estimate how much information the data provider is actually making available. To do this, we note that for estimating population level inferences a differentially private data access system, with our algorithm implemented, is equivalent to an ordinary data access system with some specific proportion of the data discarded. Indeed, it turns out we can calculate *the proportion of observations effectively lost due to the privacy protective procedures*, after one run of our algorithm without any addition expenditure from the privacy budget. We recommend that the estimates we now offer be made publicly available by all data providers or researchers using differentially private data access systems.

To make this calculation, define $\hat{\theta}_N$ as the estimator we would calculate if the data were not private and $\tilde{\theta}_N^{\text{dp}}$ as our estimator — each based on the number of observations indicated in the subscript. Then we set as our goal estimating N^* (with $N^* < N$) such that $V(\hat{\theta}_{N^*}) = V(\tilde{\theta}_N^{\text{dp}})$. Because $V(\hat{\theta}_{N^*}) \propto 1/N^*$ and $V(\tilde{\theta}_N^{\text{dp}}) \propto 1/N$, we can write $V(\hat{\theta}_{N^*}) = N \cdot V(\hat{\theta}_N)/N^* = V(\tilde{\theta}_N^{\text{dp}})$. We then write the proportionate (effective) loss in observations due to the privacy protective procedures L as

$$L = \frac{N - N^*}{N} = 1 - \frac{V(\hat{\theta}_N)}{V(\tilde{\theta}_N^{\text{dp}})}. \tag{15}$$

Finally, we can estimate the numerator of the second term as $\hat{\sigma}^2_{\text{dp}}/P$, where $\hat{\sigma}^2_{\text{dp}}$ in the numerator and the denominator are outputs from our bias correction and variance estimation algorithms (Section 4). So when a dataset has N observations, but is being provided through a differentially private mechanism, this is the equivalent to the researcher having only $LN < N$ observations and no privacy protective procedures. Since this statistic does not tax the privacy budget at all, software designers should automatically report the estimate

$$\hat{L} = 1 - \frac{\hat{\sigma}^2_{\text{dp}}/P}{V(\tilde{\theta}^{\text{dp}})} \tag{16}$$

whenever the researcher chooses to disclose $\tilde{\theta}^{\text{dp}}$.

Choosing ϵ From the point of view of the statistical researcher, ϵ directly influences the standard error of the quantity of interest, although as long as our algorithm is used this choice will not affect the degree of bias. Because we show in Section 4 that typically $\hat{V}(\tilde{\theta}^{\text{dp}}) < \hat{V}[c(\hat{\theta}^{\text{dp}}, \Lambda)] < \hat{V}(\hat{\theta}^{\text{dp}})$, we can simplify and provide some intuition by writing an upper bound on the standard error $\text{SE}_{\tilde{\theta}^{\text{dp}}} \equiv \sqrt{\hat{V}(\tilde{\theta}^{\text{dp}})}$ as

$$\text{SE}_{\tilde{\theta}^{\text{dp}}} < \sqrt{V(\hat{\theta}^{\text{dp}}) + S(\Lambda, \epsilon, \delta, P)^2}. \tag{17}$$

A researcher can use this expression to judge how much of their allocation of ϵ to assign to the next run by using their prior information about the likely value of $\hat{V}(\hat{\theta}^{\text{dp}})$ (as they would in a power calculation), plugging in the chosen values of Λ, δ and P, and then trying different values of ϵ.

Choosing Λ Although our bias correction procedure makes the particular choice of Λ less consequential, researchers with extra knowledge should use it. In particular, reducing Λ increases the chance of censoring while reducing noise, while larger values will reduce censoring but increase noise. This Heisenberg-like property is an intentional feature of differential privacy, designed to keep researchers from being able to see with too much precision.

We can however choose among the unbiased estimators our method produces that have the smallest variance. To do that, researchers should set Λ by trying to capture the point estimate of the mean. Although this cannot be done with certainty, researchers can often do this without seeing the data. For example, consider the absolute value of coefficients from any real application of logistic regression. Although technically unbounded, empirical regularities in how researchers typically scale their input variables lead to logistic regression coefficients reported in the literature rarely having absolute values above about five. Similar patterns are easy to identify across many other statistical procedures. A good software interface would thus not only include appropriate defaults but also enable users to enter asymmetric Λ intervals, as we do by reparameterization for α_2 in Section 4.1. Then the software, rather than the user, could take responsibility (in the background) for rescaling the variables as necessary.

Applied researchers are good at making choices like this as they have considerable experience with scaling variables, a task that is an essential part of most data analyses. Researchers also frequently predict the values of their quantities of interest both informally, when deciding what analysis to run next, and formally for power calculations. Statisticians and methodologists find it useful to standardize variables and use other similar techniques to enable them to ignore the substantive features of a problem, but in practice making data analyses invariant to the substance makes no sense for the researcher seeking to learn something about the world, which is their purpose for conducting the analysis in the first place. If the data surprise us, we will learn this because the $\hat{\alpha}_1^{\text{dp}}$ and $\hat{\alpha}_2^{\text{dp}}$ are disclosed as part of our procedure. If either quantity is more than about 60%, we recommend researchers adjust Λ and rerun their analysis (see Appendix B for details).

Practical Implementation Choices As with all policies, privacy policies can be informed by science but not determined by it. Policy choices are by definition always inherently political to some degree. This tension is revealed in the privacy literature by the sometimes divergent perspectives of theorists and practitioners. Theorists tend to be highly conservative in setting privacy parameters and budgets; practitioners responsible for implementing data access systems usually take a more lenient perspective. Both perspectives make sense: Theorists analyze worst case scenarios using mathematical certainty as the standard of proof, and must be ever wary of scientific adversaries hunting for loopholes in their proposed privacy protective mechanisms. This divergence even makes sense both theoretically, because privacy bounds are orders of magnitude higher than what we would expect in practice (Erlingsson, Mironov, et al., 2019), and empirically, because those responsible for implementing data access systems have little choice but to make some compromises in turning mathematical proofs into physical reality. In fact, we have been unable to find even a single large scale implementation of differential privacy that exactly meets the mathematical standard, even when theorists have been heavily involved in its design. In practice, common implementations of differential privacy allow larger values of ϵ for each run (such as in the single digits), reset the privacy budget each day, or do not have a privacy budget at all.

We thus offer here two ways of thinking about practical approaches to the use of this technology. First, although the data sharing regime can be broken by intentional attack, because we now know that re-identification from de-identified data is often possible, de-identification is still helpful — not for the ϵ bound but for reducing risk in practice. It is no surprise that university Institutional Review Boards have rephrased their regulations from "de-identified" to "not readily identifiable" rather than responding to recent discoveries by disallowing data sharing entirely. Indeed, the long history of the data sharing regime has seen exceptionally few instances where a research subject in an academic study was harmed by unauthorized re-identification (i.e., excluding computer science demonstrations), and so this technology is clearly still of considerable practical use. By adding some amount of privacy protective procedures, like noise and censoring, to de-identified data means we are further obscuring and therefore protecting private information. If the privacy budget is kept small, nothing else need be done, but if we relax this requirement and allow ϵ to be larger for any one run we will still be greatly reducing the probability of privacy violations in practice. In these situations, taking other practical steps is prudent as well such as disallowing repeated runs of the same analysis (say by returning cached results).

Second, potential data providers and regulators should ask themselves *Are these researchers trustworthy?* In the past, they almost always have been, as legitimate academics have rarely if ever used their access to sensitive data to violate the privacy of their research subjects. When this fact provides insufficient reassurance for policymakers, we can move to the data access regime. However, a middle ground does exist by trusting researchers (perhaps along with auxiliary protections, such as signed data use agreements by university employers, financial or career sanctions for violations, and full auditing of all analyses to verify compliance). With trust, researchers can be given full access to the data, be allowed to run as many analyses of whatever type they wish, but be required to use the algorithm proposed here for any results to be disclosed publicly. The data holder would then maintain a strict privacy budget summed over all published analyses, which is far more useful for scientific research than counting every exploratory data analysis run (but not necessarily disclosed publicly) against the budget. This plan may in some respects approximate the differential privacy ideal more closely than the typical data access regime, as the privacy protections among results published are then completely protected by the mathematical guarantees. There are theoretical risks (Dwork and Ullman, 2018), but the advantages to the public good that can come from research with fewer constraints may also be substantial.

Software Design We recommend data access systems that use our procedures allow a wide range of both statistical methods and quantities of interest calculated from them. Researchers should be able to choose any quantity (corresponding to θ in our algorithm) to estimate and disclose. Most of the time, researchers will wish to disclose quantities derived from statistical models rather than coefficients or other statistics typically output from statistical software. Statistical researchers view these as valuable for interpreting the data analysis, but ultimately intermediate quantities on the way to their ultimate quantity of interest. Given the limited privacy budget, researchers will want to choose which quantities to disclose much more selectively. For example, instead of logit coefficients, researchers would typically be more interested in reporting relative risks, probabilities, and risk differences (King, Tomz, and Wittenberg, 2000; King and Zeng, 2002). Even regression coefficients are often best replaced by the exact quantity of interest to the researcher, such as a predicted value, the probability that one party's candidate wins the election, or a first difference. Professional software should also allow researchers to submit statistical code to be checked and included, since the algorithm we present here can wrap around any legitimate statistical procedure.

Designing the software user interface to encourage best statistical practices can be especially valuable. This is especially so for users previously unfamiliar with differential privacy, whether they are statistically oriented or not. One simple procedure would be to always provide a simulated dataset (constructed so as not to leak any privacy from the real dataset) so users could compare the results from runs with and without the privacy protections and get a feel for how to do data analysis within a differential privacy framework.

Finally, under the topic of "do not try this at home," data providers should understand that a differentially private data access system involves details of implementation not covered here. These include issues regarding random number generators, privacy budgets, parallelization, security, authentication, and authorization. They also involve avoiding side attacks on the timing of the algorithm, statistical methods that occasionally fail (e.g., due to collinearity in regression or, in logit, perfect discrimination), the privacy budget, and the state of the computer system (e.g., Garfinkel, Abowd, and Powazek, 2018; Haeberlen, Pierce, and Narayan, 2011).

7 Concluding Remarks

The differential privacy literature focuses appropriately on the utility-privacy trade off. We propose to revise the definition of "utility" for at least some purposes so it offers value to researchers and others that seek to use confidential data to learn about the world, beyond inferences to the inaccessible private data. A scientific statement is not one that is necessarily correct, but one that comes with known statistical properties and an honest assessment of uncertainty. Utility to scholarly researchers involves inferential validity, the ability to give these informative scientific statements about populations beyond available (private or public) data. While differential privacy can guarantee privacy to individuals, researchers also need inferential validity to make a data access system safe for making proper scientific statements, for society using the results of that research, and for individuals whose privacy must be protected.

Although our goal here is a generic method, with an estimator that is approximately unbiased and applicable to a single quantity of interest at a time, more specific methods, with other properties, would be worth attention from future researchers. Inferential validity without differential privacy may mean beautiful theory without data access, but differential privacy without inferential validity may result in biased substantive conclusions that mislead researchers and society at large.

Together, approaches that are differentially private and inferentially valid may begin to convince companies, governments, and others to let researchers access their unprecedented storehouses of informative data about individuals and societies. If this happens, it will generate guarantees of privacy for individuals, scholarly results for researchers, and substantial value for society at large.

Appendix A Variance Estimation Derivations

We decompose the two variance parameters using the results following Equation 10. The first we write as $\hat{V}(\hat{\theta}^{\text{dp}}) = \hat{V}(\hat{\theta}) + S^2_{\hat{\theta}}$, where $\hat{V}(\hat{\theta})$ is the variance of the mean over P draws from a normal censored at $[-\Lambda, \Lambda]$ (divided by P), and $S^2_{\hat{\theta}}$ is the variance of the differentially private noise. The distribution from which this variance is calculated then is a three component mixture (see Equation 12). The first component is a truncated normal with mean θ_T, and bounds $[-\Lambda, \Lambda]$; the two other components are the spikes at Λ and $-\Lambda$. Begin with the following generic formula for the variance of the mean of draws from a 3-component mixture distribution with weights w_i, and component mean and variances of $E[\theta_i]$, σ_i^2 respectively:

$$V(\hat{\theta}) = \frac{1}{P} \cdot \left(\left[\sum_{i=1}^{3} w_i (E[\hat{\theta}_i]^2 + \sigma_i^2) \right] - E[\hat{\theta}]^2 \right) \tag{18}$$

with weights $\mathbf{w} = [(1 - \alpha_1 - \alpha_2), \alpha_2, \alpha_1]$, and with means for the spikes at $E[\hat{\theta}_2] = \Lambda$, and $E[\hat{\theta}_3] = -\Lambda$ and variances $\sigma_2^2 = \sigma_3^2 = 0$. Then, rearranging Equation 12, we write the truncated normal mean as

$$E[\hat{\theta}_1] \equiv \theta_T = \frac{E[\hat{\theta}] - \Lambda(\alpha_1 + \alpha_2)}{1 - \alpha_2 - \alpha_1}. \tag{19}$$

and we express the variance of the truncated normal as

$$\sigma_1^2 = \sigma^2 \left[1 + \frac{\left(\frac{-\Lambda - \theta}{\sigma}\right) Q_1 - \left(\frac{\Lambda - \theta}{\sigma}\right) Q_2}{1 - \alpha_2 - \alpha_1} - \left(\frac{Q_1 - Q_2}{1 - \alpha_2 - \alpha_1} \right)^2 \right] \tag{20}$$

where $Q_1 = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{-\Lambda-\theta}{\sigma}\right)^2\right)$ and $Q_2 = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\Lambda-\theta}{\sigma}\right)^2\right)$. σ^2 is the variance of the distribution from which partitions are drawn (before censoring).

We now use these results to fill in Equation 18:

$$V(\hat{\theta}) = \frac{1}{P} \cdot \left((1 - \alpha_2 - \alpha_1) \left(\theta_T + \sigma_1^2\right) + \Lambda^2(\alpha_2 + \alpha_1) - E[\hat{\theta}]^2 \right). \tag{21}$$

Finally, our estimator of this variance simply involves plugging in for $\{\hat{\alpha}_1, \hat{\alpha}_2, \tilde{\theta}^{\text{dp}}, \sigma^{\text{dp}}, \hat{\theta}^{\text{dp}}\}$ the values $\{\alpha_1, \alpha_2, \theta, \sigma, E[\hat{\theta}]\}$, respectively.

Next, we decompose the second parameter of the variance matrix of Equation 14 in the same way: $\hat{V}(\hat{\alpha}_2^{\text{dp}}) = V(\hat{\alpha}_2) + S^2_{\hat{\alpha}}$, the first component of which is the variance of the proportion of partitions that are censored (prior to adding noise). We represent whether a partition is censored or not by an indicator variable equal to 1 with probability α_2: If $A_p = \mathbb{1}(\hat{\theta}_p > \Lambda)$, then $\Pr(A_p = 1) = \alpha_2$. Then the sum of iid binary variables is a binomial, with variance $V\left(\sum_{p=1}^{P} A_p\right) = P\alpha_2(1-\alpha_2)$. Plugging $\hat{\alpha}_2^{\text{dp}}$ into the decomposition yields

$$\hat{V}(\hat{\alpha}) = \frac{1}{P}(1-\hat{\alpha}_2^{\text{dp}})\hat{\alpha}_2^{\text{dp}} + S_{\hat{\alpha}}^2. \tag{22}$$

Finally, we derive the covariance:

$$\begin{aligned}
\text{Cov}(\hat{\theta}^{\text{dp}}, \hat{\alpha}_2^{\text{dp}}) &= \text{Cov}(\hat{\theta}, \hat{\alpha}_2) \qquad \text{(noise is additive and independent)} \\
&= \text{Cov}\left(\frac{1}{P}\sum_{p=1}^{P} c(\hat{\theta}_p, \Lambda), \frac{1}{P}\sum_{p=1}^{P} A_p\right) \\
&= \frac{1}{P}\text{Cov}\left(c(\hat{\theta}_1, \Lambda), A_1\right) \qquad (\hat{\theta}_p \text{ and } A_p \text{ are iid over } p) \\
&= \frac{1}{P}\left\{E[c(\hat{\theta}_1, \Lambda)A_1] - E[c(\hat{\theta}_1, \Lambda)E(A_1)]\right\} \\
&= \frac{1}{P}\left\{E[c(\hat{\theta}_1, \Lambda) \mid A_1 = 1) - E[c(\hat{\theta}_1, \Lambda) \mid A_1 = 0]\right\}\alpha_2(1-\alpha_2) \qquad (23)
\end{aligned}$$

where $E[c(\hat{\theta}_1, \Lambda) \mid A_1 = 1] = \Lambda$, and $E[c(\hat{\theta}_1, \Lambda) \mid A_1 = 0] = \theta_T$, the mean of the truncated normal mean component of the censored normal. We thus use Equation 19 and plug estimates into Equation 23:

$$\text{Cov}(\hat{\theta}^{\text{dp}}, \hat{\alpha}_2^{\text{dp}}) = \frac{1}{P}\left(\Lambda - \frac{\hat{\theta}^{\text{dp}} - \hat{\alpha}_2\Lambda + \hat{\alpha}_1\Lambda}{1-\alpha_2-\alpha_1}\right)\hat{\alpha}_2(1-\hat{\alpha}_2). \tag{24}$$

Appendix B When Privacy Procedures Obscure All Relevant Information

All privacy protective procedures are designed to destroy or hide information by making it more difficult to draw certain inferences from confidential data. These are worthwhile to protect individual privacy and to ensure that data which might not otherwise be accessible at all are in fact available to researchers. However, with the noise and censoring used in differential privacy, some inferences will be so uncertain that no substantive knowledge can be learned. In even more extreme situations, our bias correction procedures, which rely on some information passing through the differential privacy filters, would have no leverage left to do their work. In this appendix, we develop a *rule of thumb* that suggests when privacy protected data analysis becomes like trying to get blood from a stone: $\max(\alpha_1, \alpha_2) > 0.6$ or $\epsilon P < 100$ (also, if $\epsilon P \gg 100$ then $\max(\alpha_1, \alpha_2)$ could be even larger before a problem occurs). If an analysis is implicated by this rule of thumb, then it is best to rerun the analysis with more partitions, use more of the privacy budget, or adjust Λ. If none of these are possible, then the only options are to negotiate with the data provider for a larger privacy budget allocation, collect more data, or abandon inquiry into this particular quantity of interest.

Recall that we attempt to choose Λ in order that each $\hat{\theta}_p \in [-\Lambda, \Lambda]$. We then keep this interval fixed and study the distribution of the mean $\hat{\theta} = \frac{1}{P}\sum_{p=1}^{P}\hat{\theta}_p$, which has a variance P times smaller than the distribution of $\hat{\theta}_p$. Now consider the unusual edge case where so much noise is added that $|\hat{\theta}^{\text{dp}}| \gg \Lambda$ (in contrast to a small deviation, which has little consequence). In this extreme situation, using $\hat{\theta}^{\text{dp}}$ as a plug-in estimator for $E(\hat{\theta}^{\text{dp}})$ no longer works because no values of θ and σ^2 can be logically consistent with it, given Λ; in some ways, such a result even nonsensically suggests that $\sigma^2 < 0$.

In this situation, we could simply stop and declare that no reasonable inference is possible and, if we do, we wind up with an analogous rule of thumb. However, to build intuition for this rule, we now show what happens if we try to accommodate this edge case computationally. Thus, if $\hat{\theta}^{\text{dp}} > \Lambda$ we learn that $e > 0$ (where e is the differentially private error defined in Equations 7-8), and so we replace $\hat{\theta}^{\text{dp}}$ with $\check{\theta}^{\text{dp}} \equiv \hat{\theta}^{\text{dp}} - S\frac{\sqrt{2}}{\sqrt{\pi}}$, where the second term is $E(e|\hat{\theta}^{\text{dp}} > \Lambda) = E(e|e > 0)$. This adjustment makes the system of equations (and the resulting $\tilde{\theta}^{\text{dp}}$) possible, at the cost of some (third order) bias. We now derive our rule of thumb by showing how to bound this bias by appropriately choosing ϵ, P, and Λ.

For simplicity, we study the dominant case of one-sided censoring ($\alpha_1 = 0$), which enables us to solve the bias correction equations algebraically rather than numerically; the results are not very different for two-sided censoring. Thus, begin with the facts, including

α_2 in Equation 11 and

$$\hat{\theta}^{\text{dp}} = (1-\hat{\alpha}_2^{\text{dp}}) \left[\theta - \frac{\frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\Lambda - \hat{\theta}^{\text{dp}}}{\sigma} \right)^2 \right)}{(1-\hat{\alpha}_2^{\text{dp}})} \right] + \hat{\alpha}_2^{\text{dp}} \Lambda. \tag{25}$$

Then solve these equations for θ, which we label $\tilde{\theta}^{\text{dp}}$ as above, and show, conditional on α_2, that $\tilde{\theta}^{\text{dp}}$ is a linear function of $\hat{\theta}^{\text{dp}}$:

$$\tilde{\theta}^{\text{dp}} = \hat{\theta}^{\text{dp}} \left(\frac{1}{B} \right) + \Lambda \left(\frac{B-1}{B} \right), \tag{26}$$

where $B = (1-\hat{\alpha}_2^{\text{dp}}) + \frac{\sqrt{2}e^{-T^2/2}}{2T\sqrt{\pi}}$ and $T = \sqrt{2} \cdot \text{erf}^{-1}[2(1-\hat{\alpha}_2^{\text{dp}}) - 1]$.

Note that if we apply our bias correction (in Section 4.1) using the exact version of $E(\hat{\theta})$ (and α_2) as an input, we would find $\tilde{\theta}^{\text{dp}} = \theta$. We are therefore interested in the discrepancy $d = E(\breve{\theta}^{\text{dp}}) - E(\hat{\theta})$, which we write as

$$\begin{aligned}
d =& \Bigg[(1 - \Pr(\hat{\theta}^{\text{dp}} > \Lambda)) \int_{-\infty}^{\Lambda} \frac{t\mathcal{N}(t|\hat{\theta}, S^2)}{(1-\Pr(\hat{\theta}^{\text{dp}} > \Lambda))} dt \\
&+ \Pr(\hat{\theta}^{\text{dp}} > \Lambda) \int_{\Lambda}^{\infty} \frac{\left(t - S\frac{\sqrt{2}}{\sqrt{\pi}}\right) \mathcal{N}(t|\hat{\theta}, S^2)}{\Pr(\hat{\theta}^{\text{dp}} > \Lambda)} dt \Bigg] - E[\hat{\theta}] \\
=& \left[E[\hat{\theta}^{\text{dp}}] - \Pr(\hat{\theta}^{\text{dp}} > \Lambda) S \frac{\sqrt{2}}{\sqrt{\pi}} \right] - E[\hat{\theta}] \\
=& - S \frac{\sqrt{2}}{\sqrt{\pi}} \times \Pr(\hat{\theta}^{\text{dp}} > \Lambda) \\
=& - \frac{2\Lambda\sqrt{2\ln(1.25/\delta)}}{\epsilon P} \frac{\sqrt{2}}{\sqrt{\pi}} \times \Pr(\hat{\theta}^{\text{dp}} > \Lambda),
\end{aligned} \tag{27}$$

where $\Pr(\hat{\theta}^{\text{dp}} > \Lambda) = \int_{\Lambda}^{\infty} \mathcal{N}(t|\hat{\theta}, S^2) dt$ has a maximum value of 0.5. As a result, the maximum value of the discrepancy is

$$\max(d) = -\frac{2\Lambda\sqrt{\ln(1.25/\delta)/\pi}}{\epsilon P}. \tag{28}$$

Making use of Equation 26, we write the maximum possible bias in $\tilde{\theta}^{\text{dp}}$ as a function of the maximum possible bias in $\breve{\theta}^{\text{dp}}$. Thus,

$$E[\tilde{\theta}^{\text{dp}}] - \theta \leq \left(\frac{1}{B} \right) \cdot \max(d) \tag{29}$$

which shows that the bias depends on $1/B$, which itself is a deterministic function of α_2.

As shown in Figure 3, which plots this relationship, if censoring (plotted horizontally) is 0.5, then $\tilde{\theta}^{\text{dp}}$ is unbiased. We also see that we can control the maximum value of $1/B$ by controlling the level of censoring. If we follow our rule of thumb and disallow censoring over 60%, then $\max_{0 \leq \alpha_2 \leq 0.6} |\frac{1}{B}| = 1$.

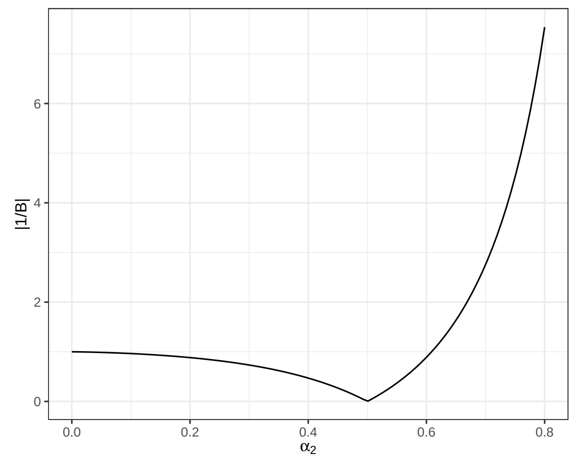

Figure 3: Relationship between $|1/B|$ and Percent Censored

To find the maximum bias under this decision rule, note that if $\Pr(\hat{\theta}^{\text{dp}} > \Lambda)$ is at its maximum, then $\alpha_2 = 0.5$ and $1/B \to 0$. It follows that $\frac{1}{B}\Pr(\hat{\theta}^{\text{dp}} > \Lambda)$ is strictly less than 0.5 and we are able to bound the absolute value of the discrepancy:

$$|E(\tilde{\theta}^{\text{dp}}) - \theta| < \left| \frac{2\Lambda\sqrt{\ln(1.25/\delta)/\pi}}{\epsilon P} \right|. \tag{30}$$

We use this result to show that we have approximately bounded the bias in $\tilde{\theta}^{\text{dp}}$ (if the computational fix is applied) relative to our quantity of interest θ. Since users set Λ on the scale of their quantity of interest to the range $[-\Lambda, \Lambda]$, the maximum proportionate bias is less than approximately

$$\frac{1}{\Lambda}\left| \frac{2\Lambda\sqrt{\ln(1.25/\delta)\pi}}{\epsilon P} \right| = \left| \frac{2\sqrt{\ln(1.25/\delta)\pi}}{\epsilon P} \right|. \tag{31}$$

For example, if we choose, from our rule of thumb, $\epsilon P = 100$ and $\delta = 0.01$, then this evaluates to 0.03, a small proportionate bias. Of course, this is the upper bound; the actual bias is likely to be a good deal smaller than even this small bound in most applications.

References

Balle, Borja and Yu-Xiang Wang (2018): "Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising". In: *arXiv preprint arXiv:1805.06530*.

Barrientos, Andrés F., Jerome Reiter, Machanavajjhala Ashwin, and Yan Chen (July 2019): "Differentially Private Significance Tests for Regression Coefficients". In: *Journal of Computational and Graphical Statistics*, pp. 1–24.

Blackwell, Matthew, James Honaker, and Gary King (2017): "A Unified Approach to Measurement Error and Missing Data: Overview". In: *Sociological Methods and Research*, no. 3, vol. 46, pp. 303–341.

Blair, Graeme, Kosuke Imai, and Yang-Yang Zhou (2015): "Design and analysis of the randomized response technique". In: *Journal of the American Statistical Association*, no. 511, vol. 110, pp. 1304–1319.

Bun, Mark and Thomas Steinke (2016): "Concentrated differential privacy: Simplifications, extensions, and lower bounds". In: *Theory of Cryptography Conference*. Springer, pp. 635–658.

Carlini, Nicholas, Chang Liu, Úlfar Erlingsson, Jernej Kos, and Dawn Song (2019): "The Secret Sharer: Evaluating and testing unintended memorization in neural networks". In: *28th {USENIX} Security Symposium ({USENIX} Security 19)*, pp. 267–284.

Desfontaines, Damien and Balázs Pejó (2019): "SoK: Differential Privacies". In: *CoRR*, vol. abs/1906.01337. arXiv: 1906.01337. URL: http://arxiv.org/abs/1906.01337.

Ding, Bolin, Janardhan Kulkarni, and Sergey Yekhanin (2017): "Collecting telemetry data privately". In: *Advances in Neural Information Processing Systems*, pp. 3571–3580.

Dwork, Cynthia, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Roth (2015): "The reusable holdout: Preserving validity in adaptive data analysis". In: *Science*, no. 6248, vol. 349, pp. 636–638.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith (2006): "Calibrating noise to sensitivity in private data analysis". In: *Theory of cryptography conference*. Springer, pp. 265–284.

Dwork, Cynthia and Aaron Roth (2014): "The algorithmic foundations of differential privacy". In: *Foundations and Trends in Theoretical Computer Science*, no. 3–4, vol. 9, pp. 211–407.

Dwork, Cynthia and Jonathan Ullman (2018): "The fienberg problem: How to allow human interactive data analysis in the age of differential privacy". In: *Journal of Privacy and Confidentiality*, no. 1, vol. 8.

Erlingsson, Úlfar, Ilya Mironov, Ananth Raghunathan, and Shuang Song (2019): "That which we call private". In: *arXiv preprint arXiv:1908.03566*.

Erlingsson, Úlfar, Vasyl Pihur, and Aleksandra Korolova (2014): "RAPPOR: Randomized aggregatable privacy-preserving ordinal response". In: *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. ACM, pp. 1054–1067.

Gaboardi, Marco, Hyun-Woo Lim, Ryan M Rogers, and Salil P Vadhan (2016): "Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing". In: *ICML'16 Proceedings of the 33rd International Conference on International Conference on Machine Learning-Volume 48*. JMLR.

Garfinkel, Simson L, John M Abowd, and Sarah Powazek (2018): “Issues encountered deploying differential privacy”. In: *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*. ACM, pp. 133–137.

Glynn, Adam N (2013): “What can we learn with statistical truth serum? Design and analysis of the list experiment”. In: *Public Opinion Quarterly*, no. S1, vol. 77, pp. 159–172.

Guess, Andrew, Jonathan Nagler, and Joshua Tucker (2019): “Less than you think: Prevalence and predictors of fake news dissemination on Facebook”. In: *Science advances*, no. 1, vol. 5, eaau4586.

Haeberlen, Andreas, Benjamin C Pierce, and Arjun Narayan (2011): “Differential Privacy Under Fire.” In: *USENIX Security Symposium*.

Jayaraman, Bargav and David Evans (2019): “Evaluating Differentially Private Machine Learning in Practice”. In: *28th USENIX Security Symposium (USENIX Security 19). Santa Clara, CA: USENIX Association*.

Karwa, Vishesh and Salil Vadhan (2017): “Finite sample differentially private confidence intervals”. In: *arXiv preprint arXiv:1711.03908*.

King, Gary, Robert O. Keohane, and Sidney Verba (1994): *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press. URL: bit.ly/gKroKsV.

King, Gary and Nathaniel Persily (In press): “A New Model for Industry-Academic Partnerships”. In: *PS: Political Science and Politics*. URL: GaryKing.org/partnerships.

King, Gary, Michael Tomz, and Jason Wittenberg (Apr. 2000): “Making the Most of Statistical Analyses: Improving Interpretation and Presentation”. In: *American Journal of Political Science*, no. 2, vol. 44, pp. 341–355. URL: bit.ly/makemost.

King, Gary and Langche Zeng (2002): “Estimating Risk and Rate Levels, Ratios, and Differences in Case-Control Studies”. In: *Statistics in Medicine*, vol. 21, pp. 1409–1427. URL: bit.ly/estrrCC.

Kleiner, Ariel, Ameet Talwalkar, Purnamrita Sarkar, and Michael I Jordan (2014): “A scalable bootstrap for massive data”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, no. 4, vol. 76, pp. 795–816.

Mironov, Ilya (2017): “Rényi differential privacy”. In: *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. IEEE, pp. 263–275.

Mohan, Prashanth, Abhradeep Thakurta, Elaine Shi, Dawn Song, and David Culler (2012): “GUPT: privacy preserving data analysis made easy”. In: *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, pp. 349–360.

Monogan, James E (2015): “Research preregistration in political science: The case, counterarguments, and a response to critiques”. In: *PS: Political Science & Politics*, no. 3, vol. 48, pp. 425–429.

Nissim, Kobbi, Sofya Raskhodnikova, and Adam Smith (2007): “Smooth sensitivity and sampling in private data analysis”. In: *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*. ACM, pp. 75–84.

Politis, Dimitris N, Joseph P Romano, and Michael Wolf (1999): *Subsampling*. Springer Science & Business Media.

Roberts, Margaret E (2018): *Censored: distraction and diversion inside China’s Great Firewall*. Princeton University Press.

Sheffet, Or (2017): “Differentially private ordinary least squares”. In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 3105–3114.
Smith, Adam (2011): “Privacy-preserving statistical estimation with optimal convergence rates”. In: *Proceedings of the forty-third annual ACM symposium on Theory of computing*. ACM, pp. 813–822.
Stefanski, L. A. (2000): “Measurement Error Models”. In: *Journal of the American Statistical Association*, no. 452, vol. 95, pp. 1353–1358.
Sweeney, Latanya (1997): “Weaving technology and policy together to maintain confidentiality”. In: *The Journal of Law, Medicine & Ethics*, no. 2-3, vol. 25, pp. 98–110.
Tang, Jun, Aleksandra Korolova, Xiaolong Bai, Xueqiang Wang, and Xiaofeng Wang (2017): “Privacy loss in apple’s implementation of differential privacy on macos 10.12”. In: *arXiv preprint arXiv:1709.02753*.
Vadhan, Salil (2017): “The complexity of differential privacy”. In: *Tutorials on the Foundations of Cryptography*. Springer, pp. 347–450.
Wang, Yue, Daniel Kifer, and Jaewoo Lee (2018): “Differentially Private Confidence Intervals for Empirical Risk Minimization”. In: *arXiv preprint arXiv:1804.03794*.
Wang, Yue, Jaewoo Lee, and Daniel Kifer (2015): “Differentially private hypothesis testing, revisited”. In: *arXiv preprint arXiv:1511.03376*, vol. 1.
Warner, Stanley L (1965): “Randomized response: A survey technique for eliminating evasive answer bias”. In: *Journal of the American Statistical Association*, no. 309, vol. 60, pp. 63–69.
Wasserman, Larry (2006): *All of nonparametric statistics*. Springer Science & Business Media.
— (2012): “Minimaxity, statistical thinking and differential privacy”. In: *Journal of Privacy and Confidentiality*, no. 1, vol. 4.
Williams, Oliver and Frank McSherry (2010): “Probabilistic inference and differential privacy”. In: *Advances in Neural Information Processing Systems*, pp. 2451–2459.
Wilson, Royce J, Celia Yuxin Zhang, William Lam, Damien Desfontaines, Daniel Simmons-Marengo, and Bryant Gipson (2019): “Differentially Private SQL with Bounded User Contribution”. In: *arXiv preprint arXiv:1909.01917*.
Winship, Christopher and Robert D. Mare (1992): “Models for Sample Selection Bias”. In: *Annual Review of Sociology*, vol. 18, pp. 327–50.
Wood, Alexandra, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R O’Brien, Thomas Steinke, and Salil Vadhan (2018): “Differential Privacy: A Primer for a Non-Technical Audience”. In: *Vand. J. Ent. & Tech. L.* Vol. 21, p. 209.