ANNUAL REVIEWS

Annual Review of Statistics and Its Application

Statistical Data Privacy: A Song of Privacy and Utility

Aleksandra Slavković and Jeremy Seeman

Department of Statistics, The Pennsylvania State University, University Park, Pennsylvania, USA; email: sesa@psu.edu, jhs5496@psu.edu

Annu. Rev. Stat. Appl. 2023. 10:189–218

First published as a Review in Advance on November 18, 2022

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-033121-112921

Keywords

statistical data privacy, statistical disclosure control, formal privacy, differential privacy, inference

Abstract

To quantify trade-offs between increasing demand for open data sharing and concerns about sensitive information disclosure, statistical data privacy (SDP) methodology analyzes data release mechanisms that sanitize outputs based on confidential data. Two dominant frameworks exist: statistical disclosure control (SDC) and the more recent differential privacy (DP). Despite framing differences, both SDC and DP share the same statistical problems at their core. For inference problems, either we may design optimal release mechanisms and associated estimators that satisfy bounds on disclosure risk measures, or we may adjust existing sanitized output to create new statistically valid and optimal estimators. Regardless of design or adjustment, in evaluating risk and utility, valid statistical inferences from mechanism outputs require uncertainty quantification that accounts for the effect of the sanitization mechanism that introduces bias and/or variance. In this review, we discuss the statistical foundations common to both SDC and DP, highlight major developments in SDP, and present exciting open research problems in private inference.

1. INTRODUCTION

1.1. Statistical Data Privacy

Privacy and confidentiality (which we, in this review, synonymously refer to as "privacy") are widely viewed as an essential component of free societies (Westin 1968, Cohen 2012). As large-scale data collection becomes more commonplace, threats to individual data privacy grow ever more prominent. By using published statistics of any kind, such as summary statistics, adversaries can perform reconstruction attacks, where they attempt to identify likely realizations of the individual records comprising the confidential database (for a survey of common attacks, such as those based on solving linear programming problems for count data, see, e.g., Dwork et al. 2017). These reconstructed records can then be associated to further data sources through record linkage techniques (for a review of these techniques, see, e.g., Vatsalan et al. 2013) and thus heighten their disclosure.

Despite these threats, many statisticians and data users have a limited understanding of what data privacy is and how it affects our work. Data privacy is often synonymous with boilerplate procedures required to satisfy compliance obligations (e.g., the Health Insurance Portability and Accountability Act of 1996 or institutional review boards), an inconvenience to our normal operating procedures. We may think that the solution is always anonymization, or the act of removing personally identifying information from a database. Yet statistical outputs (such as summary statistics and parameter estimates) pose threats to individual disclosure, just as their inputs do through access to confidential databases; this makes data privacy a methodological problem beyond the known failures of anonymization alone (Ohm 2009). Not all statistics reveal the same information about individuals, and negotiating to find the proper balance between privacy and utility—that is, between privacy protections and data usefulness via uncertainty quantification of multiple sources of errors, including those from the data privacy model (see **Figure 1**)—is precisely where statistics has much to offer.

Statistical data privacy (SDP) aims to develop provable and usable data privacy theory and methodology, by integrating tools from computer science and statistics to enable broad sharing of data across many different data contexts and domains where it is desired or required that individuals' identities or sensitive attributes are protected (e.g., census, health, genomic data, social networks). SDP methods need to minimize privacy loss/disclosure risk of sensitive information while at the same time preserving sufficient statistical integrity of data in order to support valid inference (i.e., maximizing data utility). Two dominant frameworks in SDP, defined by different units of analysis, make different conceptual trade-offs about adversarial assumptions (e.g., a worst-case adversary may know all but one record in the database); disclosure risks, which define particular ways of quantifying privacy harms from disclosure (e.g., quantifying how much information the adversary learned from the published statistics); and their effects on downstream inference. Statistical disclosure control (SDC) and statistical disclosure limitation (SDL) methods (e.g., Hundepool et al. 2012) typically analyze individual databases, whereas differential privacy (DP) methods (Dwork & Roth 2014) analyze pairs of databases in a shared schema (space of possible databases). The terms SDC and SDL are often used interchangeably.

Despite good-faith attempts to unify these perspectives, some dating back to the onset of DP, as discussed by Slavkovic (2013), SDC and DP research perspectives still diverge. SDP scholars, data administrators, and quantitative social scientists, it seems, have become more polarized as proponents of one perspective or the other. In particular, the US Census Bureau's decision to use DP has led to debates about the merits of either approach (Abowd 2021, Domingo-Ferrer et al. 2021). Furthermore, this insularity has led to growing gaps between theoretical developments, now dominated by DP research, and applied methodology, now dominated by SDC research. Such debates

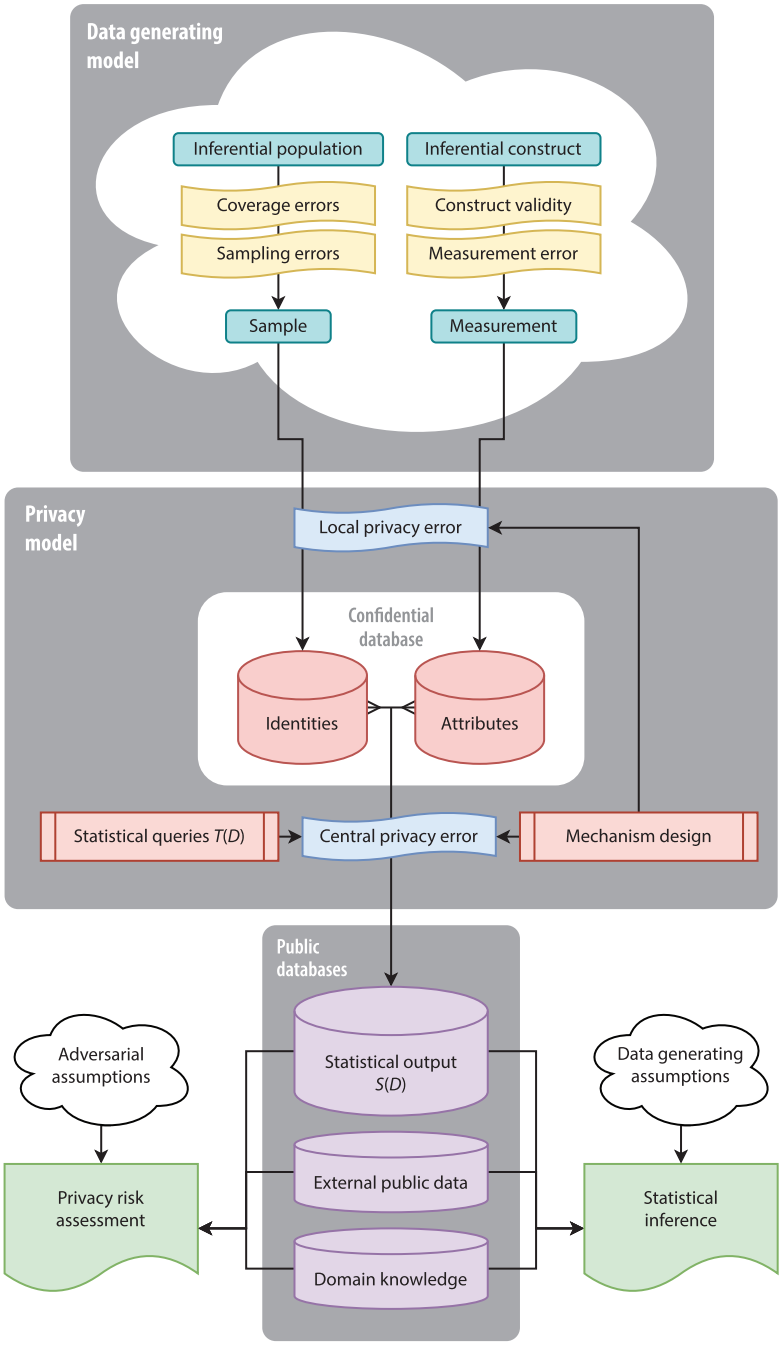

Figure 1

Flowchart of statistical data privacy modeling for privacy risk and data utility assessment.

remain important, as they highlight the inherently political nature of privacy-preserving data stewardship (Rogaway 2015, boyd & Sarathy 2022) and the way mathematical formalisms frame discussions about privacy (Seeman & Susser 2022). Still, SDC and DP share more similarities than differences, and unifying these ideas can help mutually enrich future SDP research.

In this article, we have three main goals: First, we highlight how SDC and DP methods are built upon common statistical foundations that make different but necessary compromises in conceptualizing privacy. Second, we discuss how SDP is inseparable from the study of data generating processes in both designing optimal private estimators and adjusting inferences for privacy preservation given sanitized outputs. Third, we showcase open statistical problems typically left unarticulated by theoretical SDP research, such as valid statistical inference, computational tractability, and compatibility with probability models. The article is organized as follows: Sections 2 and 3 review privacy quantification and release mechanisms/methods using SDC and DP, respectively. Section 4 reviews inferential problems common to all SDP. Section 5 presents a fabulistic case study applying SDP methodology to a small population of the Westerosis, fictional characters from the *Song of Ice and Fire* book series (e.g., Martin 2011) and the television series *Game of Thrones*. Finally, Section 6 discusses current research practices and proposes future research directions to improve SDP research.

Schema: the space of all possible confidential databases, $\mathcal{D}$

Output space: $\mathcal{S}$, the space of all possible statistical outputs

Release mechanism (RM): a randomized function S from $\mathcal{D}$ to $\mathcal{S}$ that sanitizes a database D's output

Sanitized output: $S(D)$, a statistical output from a confidential database that has undergone some form of privacy preservation

Microdata: record-level data, either directly from the confidential database, or synthetically generated from a model

1.2. Notation and Problem Formulation

Throughout this article, we use the following notation and terminology. Let $\mathcal{X}$ be the sample space for one individual's contributions to a database, and let $\mathcal{D} = \mathcal{X}^n$ be the sample space for a confidential database of n individuals who have contributed their data. We refer to $\mathcal{D}$ as the schema, or the space of possible databases. For the purposes of this review, we assume the unit of observation refers to one individual; however, similar methodology may be applied to groups, businesses, or organizations.

In SDP, we release a statistic from a set of possible statistical outputs, which we call the output space, $\mathcal{S}$. A release mechanism (RM), defined by $S : \mathcal{D} \mapsto \mathcal{S}$, is a transformation of the confidential data that produces the sanitized output, $S(D)$. SDP characterizes the privacy risk and data utility properties of different RMs in different scenarios. RMs may be either deterministic, i.e., $S(D)$ transforms the confidential database D according to a fixed function such as aggregation, or randomized, i.e., $S(D)$ is a random variable that varies conditionally on D, such as a statistic with randomized noise. Moreover, $\mathcal{S}$ can take many forms, ranging from synthetic or tabular microdata releases to summary statistics and model parameter estimates. SDC and DP typically ask two different questions about the relationship between the RM S and the database $D \in \mathcal{D}$ with respect to disclosure risk:

1. SDC perspective: How does a particular statistical release $S(D)$ for a particular database $D \in \mathcal{D}$ limit a particular measure of disclosure risk, with respect to an individual or subpopulation, dependent on existing adversary knowledge?
2. DP perspective: How can a statistical RM S limit the ability for adversaries to distinguish between similar databases $D, D' \in \mathcal{D}$, such as by only changing one individual's contribution, within the same schema $\mathcal{D}$?

In asking these questions, SDC and DP frame the problem of data privacy with different conceptual compromises. SDC defines a narrow set of adversarial contexts with the goal of quantifying a risk measure on a particular database, which we refer to as an absolute measure of disclosure risk. Alternatively, DP defines a broader set of adversarial contexts with the goal

of quantifying a measure of disclosure risk differences for any two similar databases within the schema, which we refer to as a relative measure of disclosure risk.

SDC and DP frameworks are both on a spectrum of many possible ways to reason about SDP, but both require negotiating between privacy and utility. As the number of statistics released about any database increases, one can increasingly reconstruct individual records contained in that database (Dinur & Nissim 2003) (this problem, in part, motivated the start of DP research). At the same time, assumptions about data generation processes and context-specificity are necessary to provide meaningful data utility for any such statistical data release (a.k.a., "no free lunch in data privacy"; Kifer & Machanavajjhala 2011). Therefore, it behooves us to investigate both SDC and DP simultaneously.

For historical context, SDC methods have been used in official statistics since the 1960s (McKenna 2019). Early attempts to mathematically formalize SDP date back to the seminal works of Tore Dalenius (1977, p. 433), who viewed privacy as striving to reduce the harms of population-level statistical inferences on individuals:

> If the release of the statistics $T(D)$ makes it possible to determine the value [of confidential statistical data] more accurately than is possible without access to $T(D)$, a disclosure has taken place.

However, modern literature in both SDC and DP has shown that we cannot learn anything about a population without also learning something about individuals within that population (Dwork & Naor 2010), and thus, the above definition is unobtainable. As an oft-cited counterexample from Dwork et al. (2017, p. 78),

> Releasing the fact that smoking and lung cancer are strongly correlated reveals sensitive information about any individual known to smoke; however, we do not consider this to be a privacy violation, as learning this correlation has nothing to do with the use of that individual's data.

Distinctions between the statistical population-level inference and the inference about the specific individuals in the sample have become a point of confusion due to misinterpretation of DP guarantees (Kenny et al. 2021), leading, for example, Ullman (2021) to recently clarify why statistical inference is not a privacy violation. In our Fable of the Westerosi Census (Section 5), Littlefinger may learn that Dornish folk are more likely to survive than the Free folk, and thus the Dornish princess is as well, but he would have inferred that without knowing her data per se. Here, we further clarify that such a feature is central to the entire project of SDP, not solely DP.

So, how does SDP become part of the broader project of statistical inference? **Figure 1** graphically illustrates where privacy preservation sits in statistical inference, with a special focus on the social science context. Because we are working with human data, we are prone to many error sources, as often systematized by survey methodology (Groves et al. 2011). These errors influence the data generating process leading to the creation of our confidential database records, even before any sanitization is introduced.

Note that we have two kinds of errors due to privacy, which enter the data generating process at different stages and correspond to different database trust models (Stoller 2011). In the local model, privacy protecting errors are introduced into the way users contribute to the confidential database (Evfimievski et al. 2003). Alternatively, in the central model (Dwork & Roth 2014), user contributions are combined and transformed into sanitized releases. Local models confer stronger privacy guarantees because, unlike the central model, in the local model certain information about users is inaccessible even to the data curator.

We argue that the statistical perspective is essential because both risk assessment and inference on statistical output rely on the probabilistic transformations throughout the data generating process that influence said statistical output. Much of the focus on SDP research narrowly considers

only the privacy model, i.e., only the relationships between the confidential database and the statistical outputs. By taking a bird's-eye view of this process in **Figure 1**, we see many possible avenues for statistically motivated privacy research. These research avenues depend on where we as statisticians are involved in the process. Namely, do we have any say in choosing the RM? Depending on the answer, we can consider two different broad classes of problems.

Disclosure risk measure (DRM): a function R from $\mathcal{S}$ to $\mathcal{R}$ that quantifies the disclosure risk of a sanitized output, $R(S(D))$

1. Design problems: If $\mathcal{Q}$ is a class of RMs with the same privacy guarantees, how do we find an optimal RM $S^* \in \mathcal{Q}$ and associated optimal estimator $\hat{\theta}^*(S^*(D))$ for some $\theta \in \Theta$, and what is the uncertainty in $\hat{\theta}^*$?
2. Adjustment problems: Given a sanitized statistical result $S(D)$, how do we find an optimal estimator $\hat{\theta}^*(S(D))$ for some $\theta \in \Theta$, and what is the uncertainty in $\hat{\theta}^*$?

Each approach requires a different means of uncertainty quantification, as we have more flexibility when we get to decide the form of S. Still, both problem classes are equally important, since we as statisticians may be working directly with confidential data, or we may be working with private synthetic data.

2. STATISTICAL DISCLOSURE CONTROL

SDC operationalizes the trade-off between risk and utility within the context of a single observed database $D \in \mathcal{D}$. Data curators construct RMs $S(D)$ and analyze their privacy properties using a disclosure risk measure (DRM), $R : \mathcal{S} \mapsto \mathcal{R}$. The RM S is then altered depending on both the data utility offered by $S(D)$ compared with D, and the risk $R(S(D))$ compared with $R(D)$. In doing so, the confidential data are reused by the data curator multiple times in order to calibrate the balance between privacy and data utility.

2.1. Statistical Disclosure Control Methods

As originally formulated by Duncan & Pearson (1991), SDC methods for matrix-valued databases often belong to a class of linear transformations, i.e., $S(D) \triangleq ADB + C$. Here, A describes a record-level transformation, B describes variable-level transformations, and C describes additional displacement or randomized noise. Note that in practice, these transformations need not be linear, nor randomized. Regardless, each of these transformation classes introduces bias and variance to the confidential data, and thus new considerations into the data analysis process, which we briefly review:

- Record-level transformations include members of a confidential database in the released statistics with varying probabilities. Common approaches involve random sampling, outlier removal, and special unique removal (Willenborg & De Waal 1996).
- Variable transformations typically shrink $\mathcal{S}$ relative to $\mathcal{D}$. For tabular data, cell suppression, recoding, top-coding, and aggregation are all examples of schema reduction techniques that reduce $\mathcal{S}$ (Hundepool et al. 2012).
- Randomized masking injects randomized noise into quantitative statistics to prevent direct inference on any statistic exactly calculated from individual records. Note that even though randomization is a central component of DP, randomized SDC methods date back to the 1960s with randomized response (Warner 1965).

As a generalization of the approach of Duncan & Pearson (1991), synthetic data generation methods produce sanitized output similar to D but with records randomly sampled according to some model, which may be parametric or nonparametric. This sampling can occur for part of any individual record or for the entire record, and we could produce partially or fully synthetic data.

More details are provided by Drechsler & Reiter (2010), Snoke et al. (2018b), and a recent review article by Raghunathan (2021).

2.2. Disclosure Risk Measures

Given these methods, we now turn to what $R(S(D))$ means in practice. The DRM R captures information about individuals that can be inferred from the statistical release $S(D)$. Choices for R make different implicit and explicit assumptions about what adversaries know in advance, what constitutes statistical disclosure, and how to quantify the probability of those potential disclosures. In general, there are three broad categories of DRMs:

1. Quasi-identifiability measures: Quasi-identifiability is the ability for combinations of certain covariates to isolate individuals in the dataset; e.g., S satisfies k-anonymity (Sweeney 2002) if

$$R(S(D)) = \min_{X \in \mathcal{X}} \#\{i \in [n] \mid X_i = X\} \geq k, \tag{1.}$$

   where the k represents the minimum number of individuals in the database that have indistinguishable records. Notable variants include ℓ-diversity (Machanavajjhala et al. 2007) and t-closeness (Li et al. 2007), which extend k-anonymity to capture the heterogeneity of sensitive user contributions within quasi-identifying categories. For databases with nondiscrete entries, alternative approaches may be used based on clustering, microaggregation, or outlier detection (Domingo-Ferrer & Mateo-Sanz 2002).
2. Model-based reidentification measures: SDC methods often involve modeling whether particular entries are reidentifiable under various modeling assumptions in the worst-case scenario (but still within the context of a single database, unlike DP). We define this event as $r_i = 1$ for $i \in [n]$ based on a probability model P_θ for $\theta \in \Theta$. This allows us to construct reidentification rates of the form

$$R(S(D)) = \frac{1}{n} \sum_{i=1}^{n} \sup_{\theta \in \Theta} \mathrm{P}_\theta(r_i = 1 \mid S(D)). \tag{2.}$$

   The effectiveness of the measure depends on the model accuracy for the individual reidentification probabilities, $\mathrm{P}(r_i = 1 \mid S(D))$. When $\mathcal{X} = k$, different techniques can be used to model the joint distribution of frequencies for categories in the population and sample (Franconi & Polettini 2004); for example, these can be based on log-linear models (Fienberg & Steele 1998, Skinner & Shlomo 2008) or survey estimation techniques (Skinner 2009). Equation 2 calculates an average reidentification rate; we may be interested in other summary statistics of $\mathrm{P}(r_i = 1 \mid S(D))$, such as their maximum in a worst-case analysis.
3. Data-based reidentification measures: While theoretical models can upper bound DRMs, we can, alternatively, lower bound these risks by attempting such database reconstruction attacks with external data sources (Domingo-Ferrer & Torra 2003, Winkler 2004). The DRM is then a linkage rate, or

$$R(S(D)) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\text{Record } X_i \in D \text{ successfully linked to a record } Z_j \in Z\}. \tag{3.}$$

   Such an approach depends on multiple factors: How are potential records $X_i \in D$ extracted from $S(D)$? What determines a successfully linked record? And how do the external data Z relate to the population? Such questions are answered by fundamental connections between database reconstruction and record linkage (Dobra et al. 2009, Vatsalan et al. 2013, Garfinkel et al. 2019).

In the current data landscape, however, there are systemic downsides to using SDC that could be viewed as weaknesses of the framework. First, data curators often cannot disclose the mathematical form of the RM S without leaking additional confidential information [e.g., see Drechsler & Reiter (2010) on data swapping, or Slavković (2004) for cell suppression, which also shows how lack of transparency negatively impacts statistical inference]. Next, SDC methods are not robust to postprocessing: there could be transformations of our releases h where $h(S(D))$ and $S(D)$ have different DRMs. Finally, SDC methods do not easily compose, in that if we have two release strategies $S_1(D)$ and $S_2(D)$ and we know their risks, it may be difficult to quantify the risk of the joint release $(S_1(D), S_2(D))$.

Privacy loss budget (PLB): a scalar parameter that quantifies DP guarantees, with smaller values conferring stronger privacy (e.g., ϵ in ϵ-DP)

3. DIFFERENTIAL PRIVACY

DP is a framework that mathematically formalizes the privacy properties of data release strategies and addresses the above shortcomings. By starting with a privacy definition and necessitating additional randomness, DP methods are provably consistent with the privacy definition and able to satisfy these three properties:

1. Methodologically transparent: Knowledge of S preserves $S(D)$'s privacy loss.
2. Robust to postprocessing: $h(S(D))$ has, at most, the same privacy risk as $S(D)$.
3. Composable: We can analytically express the privacy risk of two DP releases $S_1(D)$ and $S_2(D)$ when jointly released.

3.1. General Setup

DP was first introduced by Dwork et al. (2006b), who defined the concept of an ϵ-indistinguishable RM, now commonly known as ϵ-DP or pure DP RM. Since then, DP as a framework has spawned a massive number of new privacy definitions (Desfontaines & Pejó 2022). Because of this, it is often unclear and debated what makes any particular property emblematic of "DP." Here, we restrict ourselves to the most common privacy definitions and properties associated with the majority of DP implementations (Dwork & Roth 2014).

DP methods aim to limit the probabilistic influence of any individual's database contribution on sanitized outputs $S(D)$ [again, $S(D)$ could be any summary statistic, parameter estimation, or synthetic microdata sample]. Under very mild regularity conditions, DP methods have semantic interpretations that limit what can be inferred about an individual's contribution to an output, regardless of whether they contribute to the database or not (Kasiviswanathan & Smith 2014, Kifer & Machanavajjhala 2014). Ideally, whether individuals choose to contribute to a database should not substantively change the overall statistical results. This demonstrates a close connection between DP and robust statistics (Dwork & Lei 2009, Avella-Medina 2021, Slavkovic & Molinari 2021).

Formally, let D and D' be two databases. Let d_H be the Hamming distance between the databases, i.e., the number of elements in the databases that differ:

$$d_H(D, D') = \#\{i \in [n] \mid D_i \neq D'_i\}. \quad 4.$$

For d_H, we say D and D' are adjacent if $d_H(D, D') = 1$. We refer to this case as bounded DP, but we may alternatively consider statistics on databases whose sizes differ by 1 (known as unbounded DP).

The overall goal of DP is to ensure $S(D)$ and $S(D')$ are close together with high probability when D and D' are adjacent. This is done by parameterizing the distance between $S(D)$ and $S(D')$ with functions of scalar parameters known as privacy loss budgets (PLBs), which leads to different DP definitions. Typically, PLBs are positive, real-valued numbers that capture the trade-off

between privacy and data utility; as PLBs increase, more informative statistical results may be released with weaker privacy guarantees. Different definitions have different PLB accounting systems [e.g., ϵ for ϵ-DP, ρ for ρ-zero-concentrated DP (ρ-zCDP)]:

- ϵ-DP (Dwork et al. 2006b): Define the log max-divergence as

$$D_\infty(S(D) \,||\, S(D')) \triangleq \sup_{B \in \mathcal{F}} \log\left(\frac{\mathrm{P}(S(D) \in B)}{\mathrm{P}(S(D') \in B)}\right). \quad 5.$$

  If S satisfies ϵ-DP for some PLB $\epsilon \in [0, \infty)$, then for all $B \in \mathcal{F}$ and databases, D, D' with $d_H(D, D') = 1$:

$$\mathrm{P}(S(D) \in B) \leq \mathrm{P}(S(D') \in B)e^\epsilon. \quad 6.$$

  This is equivalent to bounding $D_\infty(S(D)||S(D')) \leq \epsilon$ for all adjacent D, D'.
- (ϵ, δ)-DP (Dwork et al. 2006a): Similarly, we can relax ϵ-DP by incorporating a relaxation parameter, $\delta \in [0, 1)$:

$$\mathrm{P}(S(D) \in B) \leq \mathrm{P}(S(D') \in B)e^\epsilon + \delta. \quad 7.$$

- ρ-zCDP (Bun & Steinke 2016): Define the Rényi divergence:

$$D_\alpha(S(D) \,||\, S(D')) \triangleq \frac{1}{\alpha - 1} \int \log\left(\frac{p(S(D))^\alpha}{p(S(D'))^{\alpha-1}}\right) d\mathcal{S}, \quad 8.$$

  where $p(\cdot)$ is the density of the mechanism, and the integral is taken over the statistical output space $\mathcal{S}$. Then an RM satisfies ρ-zCDP if, for all $\alpha \in (1, \infty)$,

$$D_\alpha(S(D) \,||\, S(D')) \leq \alpha\rho. \quad 9.$$

Next, we present some of the nice statistical interpretations of ϵ-DP, the strongest of the three definitions above (in that satisfying ϵ-DP implies satisfying the other two definitions). In the same setup, consider the following hypotheses:

$$H_0 : X_1 = x_0, \quad X_1 \neq x_1, \quad \text{and} \quad x_0, x_1 \in \mathcal{X},$$

where we assume database rows are exchangeable, i.e., any user's contribution may serve as X_1. Wasserman & Zhou (2010) show that if $S(D)$ is an ϵ-DP result, then any procedure for testing H_0 based on $S(D)$ with type I error α has power bounded above by αe^ϵ. Note that similar hypothesis testing interpretations can be derived for other frameworks, such as f-DP (Dong et al. 2019) and ρ-zCDP (Bun & Steinke 2016), but for brevity, we do not discuss them here. From a Bayesian perspective, if π is any prior distribution on the above hypotheses, and if $S(D)$ satisfies ϵ-DP, one can argue that an adversary's prior odds of learning information about someone are similar to their posterior odds:

$$\frac{\mathrm{P}(H_1 \mid S(D))}{\mathrm{P}(H_0 \mid S(D))} \Big/ \frac{\pi(H_1)}{\pi(H_0)} \in [e^{-\epsilon}, e^\epsilon]. \quad 10.$$

All these interpretations capture the important property that DP only protects against relative disclosure risks; DP does not make inferences about individuals in a confidential sample impossible, but it does limit how much easier it becomes when using the sanitized statistics.

3.2. Differential Privacy Release Mechanisms

RMs that satisfy DP rely on randomization to ensure the distance between the two distributions of the output is close. Below is but a small sample of the many possible RMs used to satisfy DP. In this section, we consider properties of a statistic $T(\mathcal{D}) \in \mathcal{T}$ that we aim to release. Central to

many different DP definitions is the concept of sensitivity, defined as Δ where, for some norm on $\mathcal{S}$, $||\cdot||$:

$$\Delta \triangleq \sup_{D,D' \in \mathcal{D}, d_H(D,D')=1} ||T(D) - T(D')||.$$

Sensitivity: the largest possible change, Δ, in a statistic, $T(D)$, by altering one entry in the database D, as measured by a norm $||\cdot||$

Sensitivity captures the worst-case influence of one individual on T, which depends on $||\cdot||$. Notably, the optimal choice of norm for any particular $T(D)$ can be inferred by the geometric properties of the sensitivity space $\mathcal{S}_T$ (Awan & Slavković 2020):

$$\mathcal{S}_T \triangleq \{T(D) - T(D)' \mid D, D' \in \mathcal{D}, d_H(D, D') = 1\}.$$

For many statistics of interest, the sensitivity space (and thus Δ) is bounded by construction. As an example, count data within a single cell (i.e., the number of database users with a particular attribute) have a sensitivity of 1. However, for more complex statistics, Δ may be unbounded (e.g., a parameter in a linear regression). The most common approach to addressing this problem is to introduce enforced bounds, either by bounding the output space $\mathcal{S}$, database input space $\mathcal{D}$, or parameter space in a potential model for the data, Θ. This implementation choice has important consequences for validity and consistency of downstream statistical inference (see Section 4.4). We also note that some Bayesian sampling procedures (e.g., Wang et al. 2015, Minami et al. 2016) offer DP protections given regularity conditions on the chosen prior—i.e., there are some Bayesian models for which, if the prior is influential enough, sampling from the posterior can satisfy DP.

3.2.1. Primitive elements. Given any algorithm that accesses private data, there are many possible choices of DP mechanisms. Statistics can be sanitized by directly applying a primitive algorithm, a foundational tool that satisfies DP for a particular operation. Here we review some of these essential building blocks, and we give more specific examples in Section 4.3. More complex sanitization algorithms are created by combining these primitives or their DP outputs, as discussed in the next section.

The simplest way to satisfy DP is to add independent noise to $T(D)$, i.e., $S(D) \triangleq T(D) + \gamma$, where γ is a random variable with mean 0 and variance that increases as Δ increases and the PLB decreases. Notable examples for ϵ-DP include the Laplace mechanism (Dwork et al. 2006a), its discrete analogue (Ghosh et al. 2012), and the family of K-norm mechanisms (Hardt & Talwar 2010, Awan & Slavković 2020). Examples for (ϵ, δ) and ρ-zCDP include the Gaussian mechanism (Dwork et al. 2006a) and its discrete analogue (Canonne et al. 2020).

Alternatively, we can consider solving an optimization problem based on confidential data while simultaneously satisfying ϵ-DP. This is canonically associated with the ϵ-DP exponential mechanism (McSherry & Talwar 2007), in which a loss function $L : \mathcal{S} \times \mathcal{D} \times [0, \infty]$ is minimized while respecting ϵ-DP. When

$$\sup_{s \in \mathcal{S}} \sup_{D,D' \in \mathcal{D}} ||L(s, D) - L(s, D')|| \leq \Delta_L, \tag{11.}$$

we can satisfy ϵ-DP by releasing one sample from the density

$$f(s) \propto \exp\left(-\frac{\epsilon}{2\Delta} L(s, D)\right) \nu(s), \tag{12.}$$

where $\nu(\cdot)$ is a base measure that does not depend on D. Notable choices that allow for nice asymptotic properties include the inverse sensitivity mechanism (Asi & Duchi 2020) and K-norm gradient mechanism (Reimherr & Awan 2019), which are equivalent for some common classes of learning problems.

Some optimization problems in statistics and machine learning can be solved by perturbing the input to the problem, i.e.,

$$S(D) = \text{argmin}_{s \in \mathcal{S}} [L_D(s) + \gamma], \tag{13.}$$

where the form of γ is chosen based on the problem constraints and the PLB. These have been used in empirical risk minimization (Chaudhuri et al. 2011), convex optimization (Kifer et al. 2012), and robust M-estimation (Slavkovic & Molinari 2021), to name a few.

3.2.2. Postprocessing and composition techniques. DP's properties enable flexibility in constructing complex algorithms from primitive building blocks. First, postprocessing allows us to construct DP statistics by transforming DP microdata under the same PLB. Second, we can generate DP parameter estimates for two different models and understand their privacy guarantees using sequential composition [e.g., from the same data release $S_1(D)$ with ϵ_1 and $S_2(D)$, ϵ_2 and the total PLB will be cumulative]. Third, we can apply a DP method to different subpopulations of interest in a database and maintain the same privacy guarantees through parallel composition (McSherry 2009).

Thus, given the primitives and their properties, there are countless ways to engineer more complex DP algorithms. Here, we highlight common clusters of techniques. First, because the primitive mechanisms depend so heavily on sensitivity, artificial regularity is often induced on D to reduce this sensitivity. While this can be done using SDC techniques (truncation, discretization, clipping, etc.), more advanced methods exploit dimension reduction to effectively reduce the sensitivity of correlated statistics, such as the high-dimensional matrix mechanism for large counting query collections (McKenna et al. 2018) or private principal component analysis for linear dimension reduction (Chaudhuri et al. 2012, Awan et al. 2019).

For large datasets, subsampling provides a natural way to reduce the effective PLB for different mechanisms (often referred to as subsample-and-aggregate in DP) (Nissim et al. 2007, Li et al. 2012). For example, if $\eta * 100\%$ of a population is subsampled from an ϵ-DP result, then the resulting effective ϵ^* is $O(\eta\epsilon)$. Natural extensions of subsampling include private bagging and boosting (Dwork et al. 2010, Jordon et al. 2019).

Private synthetic data generation can be viewed as resampling from a model with parameters privately estimated from confidential data. While the regularity introduced by Bayesian priors offers some inherent privacy protections (Wang et al. 2015), other approaches involve samples privately weighted by synthetic data utility (Snoke & Slavković 2018, Vietri et al. 2020), modeling with Bayesian networks (McKenna et al. 2019), and generative adversarial network modeling (Torkzadehmahani et al. 2019). Each of these methods offers different empirical benefits (Bowen & Snoke 2019).

As an aside, private building blocks can be used to reconstruct most machine learning methods while satisfying DP. As one example, private stochastic gradient descent (Song et al. 2013) and its countless variants have allowed for the mass proliferation of DP deep learning methods (Boulemtafes et al. 2020). These methods frequently use ρ-zCDP, which has gained popularity in the machine learning community since it relies on Gaussian noise, and learning-theoretic properties of sub-Gaussian distributions form the foundations for statistical learning theory (Bousquet et al. 2003, Vershynin 2018). We direct readers to Vadhan (2017) for a review on the sample complexity of DP.

4. DATA UTILITY UNDER STATISTICAL DATA PRIVACY

In both approaches to SDP, we release sanitized statistics $S(D)$ out into the wild. What happens next? In the previous sections, we discussed the privacy properties of $S(D)$ under SDC and DP independently; now, we consider the data utility properties of arbitrary sanitized outputs $S(D)$, regardless of their privacy semantics.

"Data utility" is itself ambiguous, so we need to unpack the term. We again let $T(D) \in \mathcal{T}$ be our statistic of interest without any privacy preservation applied (i.e., our "unsanitized" or confidential

statistic). Our goal is to perform inference on a parameter $\theta \in \Theta$. In doing so, we can ask many different questions:

- Data-based utility: How close is my sanitized output $S(D)$ to the confidential output $T(D)$?
- Comparative inferential utility: How close is a sanitized estimator $\hat{\theta}(S(D))$ to a confidential estimator $\hat{\theta}(T(D))$?
- Estimator inferential utility: How is my uncertainty for θ using $\hat{\theta}(S(D))$ different from my uncertainty for θ using $\hat{\theta}(T(D))$?

Our ability to address these questions depends on whether we are designing the RM S (e.g., release a consistent and asymptotically unbiased sanitized parameter estimate) or adjusting for the effect of RM S that we did not choose (e.g., adjust the length of the confidence interval given the sanitized statistic). When we design an RM for a specific inferential task, all three should yield the same relative comparisons between estimators (i.e., if a mechanism offers better data-based utility, it also offers better estimator inferential utility). However, when we adjust for an existing RM, these utility definitions may not offer the same relative comparisons between RMs and can even be conflicting. As an example from ϵ-DP count data, the geometric mechanism (Ghosh et al. 2012) can optimize data-based utility, but it requires postprocessing that is suboptimal for estimator inferential utility on binomial data (Awan & Slavković 2018). Therefore, we need to address these two problem classes differently.

Here, we express the design and adjustment problems as two different minimax estimation problems (though we could easily pick another loss aggregating convention) (Slavković & Karwa 2019). Suppose we want to minimize some loss function $L : \Theta \times \Theta \mapsto \mathbb{R}^+$ in the worst-case scenario over a space of possible data generating distributions $\mathcal{P}$ indexed by $\mathrm{P} \in \mathcal{P}$. For any RMs $S(D)$, this requires us to think about the marginal distributions for $S(D)$ for a given data generating distribution $\mathrm{P} \in \mathcal{P}$, i.e.,

$$\mathcal{M}_S(\mathrm{P}) = \sum_{D \in \mathcal{D}} \Pr(S(D) \mid D)\ \mathrm{P}_\theta(D). \tag{14.}$$

From the design perspective, we are given a space of RMs $\mathcal{Q}$ that satisfies some privacy guarantee. Our goal is to find the optimal RM $S^* \in \mathcal{Q}$ and estimator $\hat{\theta}_{\text{Design}}(S^*(D))$ that satisfies

$$\hat{\theta}_{\text{Design}} = \arg\min_{\tilde{\theta}, S \in \mathcal{Q}} \max_{\mathrm{P} \in \mathcal{P}} \mathbb{E}_{\mathcal{M}_S(\mathrm{P})}[L(\tilde{\theta}(S(D)), \theta)]. \tag{15.}$$

This problem has been analyzed in the local DP setting (Duchi et al. 2018) and similarly in central DP (Smith 2011). Alternatively, suppose we are only given a sample $S(D)$ from an RM we did not design ourselves. Then our inference problem requires us to find the optimal adjusted estimator $\hat{\theta}_{\text{Adjust}}(S(D))$ that satisfies

$$\hat{\theta}_{\text{Adjust}} = \arg\min_{\tilde{\theta}} \max_{\mathrm{P} \in \mathcal{P}} \mathbb{E}_{\mathcal{M}_S(\mathrm{P})}[L(\tilde{\theta}(S(D)), \theta)]. \tag{16.}$$

Regardless of whether we choose S or not, statistical inference requires that we account for the transformation $S(D)$, meaning we CANNOT treat inference given $T(D)$ the same as inference given $S(D)$, as the two variables have entirely different sampling distributions; a related issue of approximating sanitized sampling distributions is discussed by Wang et al. (2018). This is true for all SDP methods, those from SDC and DP. Not only can the distribution of $S(D)|D$ introduce randomized errors due to privacy, the sample spaces of $S(D)$ and $T(D)$ can be entirely different, even for SDC methods involving no randomization. This demonstrates that the de facto practice of naively substituting $T(D)$ with $S(D)$ can produce invalid statistical inferences, with incorrect interpretations of significance, coverage, or other properties of statistical estimators (for example, for these in a network setting, see Karwa & Slavković 2016, Karwa et al. 2017).

4.1. Specific Utility and the Design Approach

First, we consider the design problem, in which our goal is to perform inference for $\theta \in \Theta$ and design a valid estimator $\hat{\theta} = S(D)$, where the RM satisfies some privacy guarantees. In the SDC literature, data utility is frequently quantified by measures that capture statistical information lost due to S (Hundepool et al. 2012). In the DP literature, the evaluation of releases from the randomized mechanisms relies on concentration inequality results to bound probabilistic distances between $S(D)$ and $T(D)$, or equivalently $\hat{\theta}(S(D))$ and $\hat{\theta}(T(D))$ (Boucheron et al. 2013). Under consistency or other oracle assumptions, these will give us estimator inferential utility measures as well.

Focusing on uncertainty quantification directly offers a few advantages. First, we can design optimal estimators based on the degree to which they specifically influence our statistical uncertainty. Examples include power and sample size analysis for experimental data (Vu & Slavkovic 2009), confidence interval width (Karwa & Vadhan 2017), the power of finite-sample hypothesis testing procedures (Awan & Slavković 2018), and asymptotically correct inference from central limit theorem approximations (Awan et al. 2019). Second, these procedures are more user friendly, as they account for uncertainty in $\hat{\theta}$ due to privacy preservation. When we strictly measure how close $\hat{\theta}(S(D))$ is to $\hat{\theta}(T(D))$, we cannot draw the same conclusions, because such a comparison does not account for other sources of error in the data generating process.

4.2. General Utility and the Adjustment Approach

Alternatively, we consider the adjustment problem, in which we must account for an RM we did not design specifically for our inferential problem. This is the setting most often associated with private synthetic microdata or collections of sanitized statistics, suggesting different kinds of utility measures for general purpose inference and inference on specific tasks (Snoke et al. 2018b, Arnold & Neunhoeffer 2020).

Importantly, different methods for generating $S(D)$ may be compatible or incompatible with different probability models for θ. For example, if we generate sanitized estimates of sufficient statistics for θ, then we would say this model is compatible with the RM because we can account for measurement error in a way that still produces asymptotically consistent statistics (for an example in Bayesian inference, see Foulds et al. 2016). However, if this is not the case, i.e., if the confidential target of our private statistics $T(D)$ is not sufficient for the model, there are certain inferences we cannot perform at all.

For inference on general purpose data, we need to characterize the likelihood of $S(D)$ given θ by integrating out the confidential data. This can be done from the frequentist perspective, i.e.,

$$\mathrm{P}_\theta(S(D)) = \sum_{\mathcal{D}} \mathrm{P}(S(D) \mid D)\ \mathrm{P}_\theta(D), \tag{17.}$$

or from the Bayesian perspective, i.e., with prior $\pi(\theta)$,

$$\mathrm{P}(\theta \mid S(D)) \propto \pi(\theta) \sum_{\mathcal{D}} \mathrm{P}(S(D) \mid D)\, \mathrm{P}(D \mid \theta). \tag{18.}$$

Because of this necessity, DP offers a particular advantage over SDC. If S satisfies DP, then the privacy mechanism is transparent (i.e., the form of measurement errors is publicly known), and the problem reduces to a classical error-in-variables problem. For common models, we can readily rely on techniques from the existing measurement error literature, such as techniques based on generalized linear models and estimating equations (Hardin & Hilbe 2002, Carroll et al. 2006, Tsiatis 2006).

Still, incorporating these errors is easier said than done, as the integration in Equations 17 and 18 can be quite computationally difficult. In some cases, connections between approximate Bayesian computation (Beaumont 2019) and inference on noisy estimates can be used for posterior inference. Fearnhead & Prangle (2012) showed that exact inference from perturbed statistics uses the same inferential sampling procedure as ABC with normal summary statistics. This allows Gong (2022) and Seeman et al. (2020) to produce valid inference from DP-sanitized statistical results. Note that different privacy mechanisms are more or less amenable to probability models, which we see in the next section.

4.3. Statistical Properties of Release Mechanisms

In this section, we compare a few different primary DP mechanisms (as discussed in Section 3.2.1) for counting queries and discuss their statistical properties; specifically, we discuss how these choices for $S(D)$ affect the ease of downstream inference through probability models. Suppose we are interested in releasing a count of events $T(D) \in \mathbb{Z}^+$, in which our sensitivity Δ is 1. We consider different ways of releasing $T(D)$ satisfying different DP formalisms:

1. Discrete Laplace: discrete Laplace mechanism for ϵ-DP (Ghosh et al. 2012):

$$S(D) = T(D) + \varepsilon, \varepsilon \sim \text{DiscreteLaplace}(\epsilon^{-1}) \tag{19.}$$

2. Discrete Gaussian: discrete Laplace mechanism for ρ-zCDP (Canonne et al. 2020):

$$S(D) = T(D) + \varepsilon, \varepsilon \sim \text{DiscreteGaussian}(\rho^{-2}) \tag{20.}$$

3. Exponential: exponential mechanism for ϵ-DP (McSherry & Talwar 2007):

$$\text{P}(S(D) = k) \propto \exp\left(-\frac{\epsilon}{2}|k - T(D)|\right) \mathbb{1}\{k \in \{0, 1, \ldots, n\}\} \tag{21.}$$

4. Randomized Response: randomized response for local ϵ-DP:

$$S(D) = \sum_{i=1}^{n} \text{RR}(X_i), \quad \begin{cases} \text{P}(\text{RR}(X_i) = X_i) = \frac{\exp(\epsilon)}{1+\exp(\epsilon)} \\ \text{P}(\text{RR}(X_i) = 1 - X_i) = \frac{1}{1+\exp(\epsilon)} \end{cases} \tag{22.}$$

Note that not all privacy guarantees are the same: Local ϵ-DP is stronger than ϵ-DP, which is stronger than ρ-zCDP. Furthermore, each mechanism has different statistical properties, which we summarize in **Table 1** and describe here:

1. Error independence: Can randomized errors due to privacy be expressed as a perturbation, where for some norm $||\cdot||$, $||S(D) - T(D)|| \perp\!\!\!\perp D$?
2. Unbiased: Does the mechanism introduce bias into the estimate of the confidential data, i.e., does $\mathbb{E}[S(D)] = T(D)$?

Table 1 Counting mechanisms and their statistical properties

Mechanism	Discrete Laplace	Discrete Gaussian	Exponential	Randomized response
Trust model	Central	Central	Central	Local
Formalism	ϵ-DP	ρ-zCDP	ϵ-DP	ϵ-LDP
Error distribution $\perp\!\!\!\perp D$?	Yes	Yes	No	No
Unbiased?	Yes	Yes	No	No
Mode unbiased?	Yes	Yes	Yes	No
Domain-constrained?	No	No	Yes	Yes

Abbreviations: CDP, concentrated differential privacy; DP, differential privacy; LDP, local differential privacy.

3. Mode unbiased: Is the maximum likelihood output of the mechanism the confidential response, i.e., is it true that

$$\max_{S^* \in \mathcal{S}} \mathrm{P}(S(D) = S^*) = T(D)? \tag{23.}$$

(Note that the result assumes the existence of a probability mass function for $S(D)$, with analogous results for arbitrary measures or densities.)

4. Domain-constrained: Does the space of $T(D)$ conform to the space of $S(D)$, i.e., is it true that for all $B \in \mathcal{F}$,

$$\mathrm{P}(T(D) \in B) = 0 \Rightarrow \mathrm{P}(S(D) \in B) = 0? \tag{24.}$$

Note that we could postprocess either discrete Laplace or discrete Gaussian to restrict the domain, i.e.,

$$S_{\text{post}}(D) = \begin{cases} 0 & T(D) + \varepsilon < 0 \\ n & T(D) + \varepsilon > n \\ T(D) + \varepsilon & \text{Otherwise.} \end{cases} \tag{25.}$$

This postprocessing transformation, proposed by Ghosh et al. (2012), offers a uniform improvement in utility as measured by the distance between $S(D)$ and $T(D)$, i.e.,

$$\mathbb{E}[|S_{\text{post}}(D) - T(D)|] \leq \mathbb{E}[|S(D) - T(D)|]. \tag{26.}$$

However, $S_{\text{post}}(D)$ is no longer unbiased, and the errors are now data dependent. This demonstrates that postprocessing changes the statistical properties of RMs, and improving utility compared with confidential results can have unintended consequences for the statistical properties of these estimators. In fact, postprocessing can degrade the power of resulting statistical inferences, sometimes uniformly (Seeman et al. 2020, 2022). Therefore, it is essential that we consider which mechanisms are amenable to downstream inference and which make it prohibitively difficult or computationally expensive.

4.4. Risk and Model Misspecification

As one last important caveat, we remind ourselves that whenever we make modeling assumptions, there is always the potential to be wrong. SDP introduces new opportunities for different kinds of misspecification that we briefly discuss here.

SDP relies on the properties of the database schema $\mathcal{D}$ being correctly specified. When this is not the case, SDP risk and utility can both suffer. Unanticipated records falling outside the expected schema could result in weaker DRMs (e.g., negative counts being dropped); if those records are systematically excluded due to processing errors, then we could be subject to an unknown form of unaccounted missing data. As another example, we may incorrectly specify the sensitivity of a statistic $T(D)$, meaning our privacy guarantees are realized at a larger PLB than intended.

SDP risk measures may also be based on implicit adversarial assumptions, which may not hold in practice. For example, Pufferfish privacy (Kifer & Machanavajjhala 2014) offers an interpretation of the change from prior adversary knowledge to posterior adversary knowledge gained from statistics released with DP. However, this interpretation only holds for a class of priors where each user contributes independently to the database. In other words, when there is dependence between database records, this interpretation may not hold (Liu et al. 2016). This has motivated methods for the private analysis of correlated data (Song et al. 2017, Karwa et al. 2017, Seeman et al. 2022). Still, this is but one example where we must be careful to not overstate the disclosure risk protections afforded by SDP.

Table 2 Data variable descriptions for *Game of Thrones* dataset

Variable	Kind	Description
`Title`	Nominal	Title (e.g., Ser, Lord) (261 levels)
`TitleReduced`	Nominal	Major title categories (11 levels)
`Culture`	Nominal	Culture (e.g., Dornish, Braavosi, Dothraki) (35 levels)
`CultureReduced`	Nominal	Major culture categories (14 levels)
`House`	Nominal	House (e.g., Stark, Lannister, Tyrell) (326 levels)
`HouseReduced`	Nominal	Major houses (27 levels)
`Gender`	Binary	Male or female
`Nobility`	Binary	Noble or peasant
`Alive`	Binary	Alive or dead

5. THE FABLE OF THE WESTEROSI CENSUS

Many of the core research problems in SDP rely on translating statistical notions into practical commitments to privacy protections and data utility goals. Here, we do not focus on the newest nor the most advanced mechanisms by modern publishing standards. Instead, the goal of this section is to showcase that there is complex interplay between different privacy formalisms, the underlying data structure, and data generating assumptions, all of which affect risk and utility—that is, valid statistical inference. To that end, we turn to a fabulistic case study.

We consider a dataset (**https://got.show/**) based on the population of fictional characters from the continent of Westeros and its surroundings in the fantasy series *Game of Thrones* (*GoT*) based on the series of novels (beginning with Martin 2011); data were gathered by mining the text of the fan-written Wiki, *A Wiki of Ice and Fire* (**https://awoiaf.westeros.org/index.php/Main_Page**). Variables are described in **Table 2**. Note that in working with this dataset, we do not intend to make light of the very real harms caused by privacy violations. Instead, we use a dataset where such harms are impossible by construction, and the worst possible harms for readers are minor story spoilers. And to the best of our knowledge, it is impossible to violate the privacy of a fictional character.

Our data curator, Lord Varys (henceforth LV), is tasked with conducting a census of the citizens of Westeros in order to count and report which citizens have survived the politically tumultuous events of *GoT*. However, he is concerned about an adversary, Littlefinger (henceforth LF), learning information about vulnerable members of the royal family whose data may be contained in the census. For this case study, we consider whether or not a character survives the events of *GoT* to be a sensitive attribute. We demonstrate how LV's assessments of risk and utility change in different scenarios.

5.1. Counting Queries with Statistical Disclosure Control

First, LV considers different contingency tables aggregated based on different quasi-identifiers, (i.e., combinations of different nominal and binary variables). For example, he could create a `Culture+Nobility` table, which lists the number of nobles and peasants from every culture. He soon realizes that too many respondents have unique combinations of titles, cultures, and/or houses; for example, there is only one Dornish princess in the data, and LF could learn attributes about the Dornish princess if the database is released as-is. Therefore, to address these issues, he creates simplified versions of all these categorical random variables, only keeping categories with at least 10 respondents (as shown in **Table 2**) and grouping all others into a separate category. We refer to these as the reduced versions of these nominal variables, e.g., `CultureReduced`.

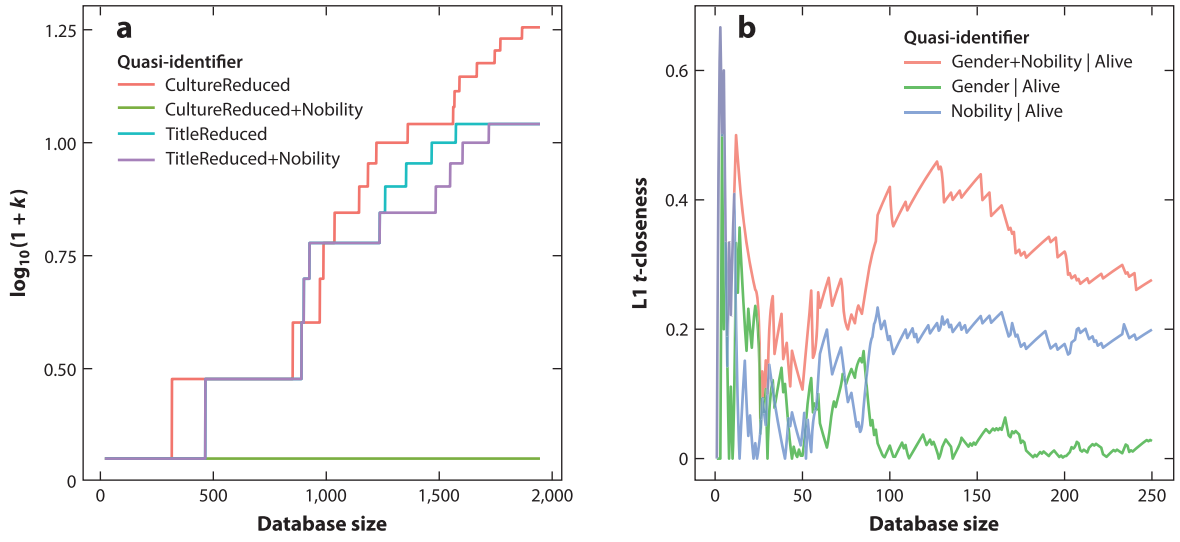

Figure 2

Game of Thrones data: (*a*) *k*-anonymity and (*b*) *t*-closeness risk measures for different aggregated counts by query (quasi-identifiers) and database size.

Next, LV considers two SDC risk measures for different contingency tables. First, he looks at *k*-anonymity (Equation 1), which captures the smallest number of census respondents with any given quasi-identifier; LV desires larger *k* values for stronger privacy protections. In **Figure 2*a***, he plots the *k*-anonymity of the first *s* rows on the *y*-axis with *s* on the *x*-axis and sees that different combinations of quasi-identifiers offer different protections; `CultureReduced` generally has the best *k*-anonymity guarantees, and these increase as the database size increases. However, LV notices that some of the cultures, even after aggregating, have only one noble, meaning that using `CultureReduced+Nobility` offers 1-anonymous privacy, regardless of the database size. This means for any database size, if LV releases `CultureReduced+Nobility`, LF will be able to reidentify anyone with a unique reduced culture and nobility status, and at least one such person exists in the database.

LV is also concerned about what LF might do if he learns which kinds of people are most affected by the events in *GoT*. So he measures *t*-closeness based on survival, or the largest difference in survival rate between any quasi-identifying group and the overall sample. In **Figure 2*b***, he sees that the *t*-closeness is small for `Gender`; that means he can release information about the survival rates for men and women in Westeros, and LF is not likely to learn much about anyone from the census's survival rate simply because of their gender. However, the *t*-closeness is larger for `Gender+Nobility`; this raises concerns for LV because he is concerned LF might learn about the probability that someone in the database, like the Dornish princess, is alive or not. However, *t*-closeness can sometimes capture population-level effects, so maybe LF would learn about the difference in survival rates between noble females and the whole population, regardless of whether the Dornish princess completed the census of Westeros or not.

LV decides he will release a 2×2 contingency table of nobility versus survival with different *k*-anonymity protections. He wants to perform inference for the null hypothesis H_0: Nobles are at least as likely as peasants to survive *GoT*, with H_1 otherwise. Under the null, the number of surviving nobles follows a hypergeometric distribution. Normally, LV would simulate from this

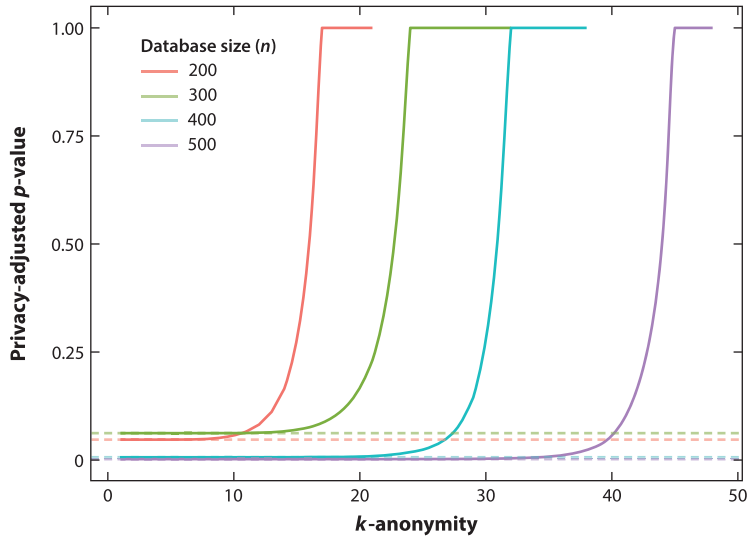

Figure 3

Game of Thrones data, privacy-corrected p-values for Fisher's exact test (H_0 = nobles at least as likely as peasants to survive) from k-anonymous tabular data at different database sizes and k values. The dashed lines represent the nonprivate p-values.

distribution and numerically estimate the test statistic, since he wants to extract as much information as possible. But LV knows his inference is affected by his choice of k, and while others may ignore that fact, he conditions on the database being k-anonymous for different k values by rejecting simulated statistics that violate k-anonymity. LV then uses these reduced samples to calculate the p-value, on the y-axis in **Figure 3**. In this risk-utility plot, the dashed horizontal lines refer to the nonprivate p-values, and the solid lines are the estimated p-values (on the y-axis) at different ks (on the x-axis). As k increases, LV loses power to detect differences in the survival rate of nobles and peasants; for example, if $n = 300$ and $k = 20$, we fail to detect such a difference at a type I error of 0.10. This demonstrates that SDC measures affect statistically valid inference, even when no randomization occurs.

5.2. Counting Queries with Differential Privacy

Because of LV's concerns, including that he cannot release much data under the SDC framework since the risk with `Nobility` is high, and the fact that he really cannot be sure what LF may already know, he decides to use bounded ϵ-DP methods for counting queries. He first plans to release the counts of the surviving and dead, aggregated by `CultureReduced+Gender+Nobility`, and with discrete Laplace noise added to each count (Equation 19). This way, LF will not be able to learn more about individuals in the database because the possible releases where any one respondent survived *GoT* or not are close together with high probability. But this has LV wondering: How much could LF know already, and how much could he stand to learn from using LV's DP census results? LV decides to analyze two different scenarios: First, maybe LF randomly guesses (i.e., flips a fair coin) to determine whether someone has survived *GoT* or not. Second, maybe LF knows the true confidential proportion of people who have survived *GoT*, in which case he is a more informed adversary (for reference, around 25% of characters in *GoT* die). Using these two priors, which we name `LessInformed` and `MoreInformed`, respectively, LV calculates the posterior risk that LF

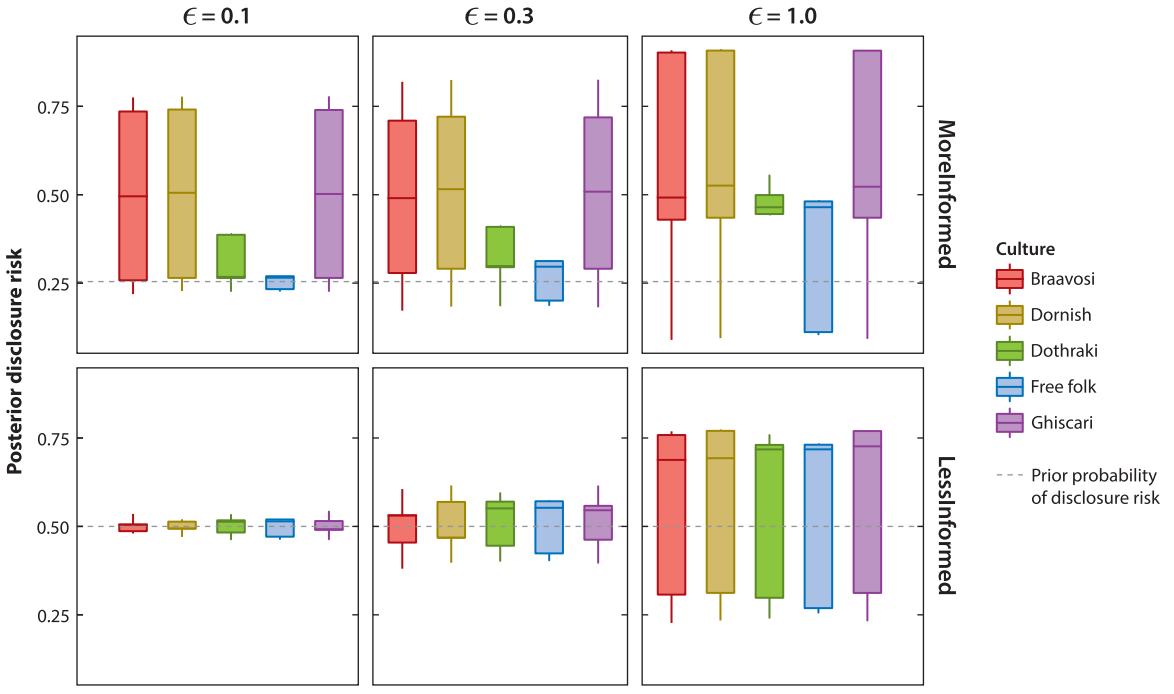

Figure 4

Box plots of posterior disclosure risks for individual *Game of Thrones* characters by prior assumptions, culture, and privacy loss budget.

learns whether the last person in any quasi-identifying cell (`CultureReduced+Gender+Nobility`) survived or died, in the worst-case scenario where LF knows all but the last entry in any table cell.

In **Figure 4**, LV plots the posterior disclosure risk on the y-axis and looks at these posterior disclosure risks. First, the worst-case disclosure risks are not the same for all cultures. Certain minority cultures under the informed prior (**Figure 4**, *top row*), such as Braavosi and Dornish, have a larger posterior probability of being reidentified within the sample at small PLBs (e.g., $\epsilon = 0.1$) than people of larger cultures, such as Free folk. Second, we see that prior assumptions can change how these posterior disclosure risks are distributed among the groups of people within the sample; if LF has a good prior, he could potentially learn more about whether the Dornish princess survives *GoT* than, say, someone from the Free folk. This visually demonstrates that even though the worst-case privacy guarantee in ϵ-DP applies to everyone, not everyone has the same posterior disclosure risks.

LV realizes from that analysis that if he wants to release the overall survival rate (or, equivalently, the death rate), he needs to sanitize it, even though it is just a summary statistic. But LV still wants to do inference on what this value could be, requiring him to account for additional noise due to privacy. In **Figure 5**, LV plots three important figures at varying PLBs, from $\epsilon = 0.01$ (stronger privacy protections) to $\epsilon = \infty$ (nonprivate release). **Figure 5*a*** shows the likelihood of survival where we assume $S(D) = T(D)$ (chosen for comparison with nonprivate inference). For this figure, LV also plots 95% confidence intervals based on the sampling distribution of $S(D)$, which increase in size as the PLB decreases. For some PLBs, like $\epsilon = 0.10$, the errors due to privacy are dominated by errors due to sampling. For other PLBs, like $\epsilon = 0.01$, the opposite is true. This essential information can only be inferred by comparing the probability model to the errors due to privacy.

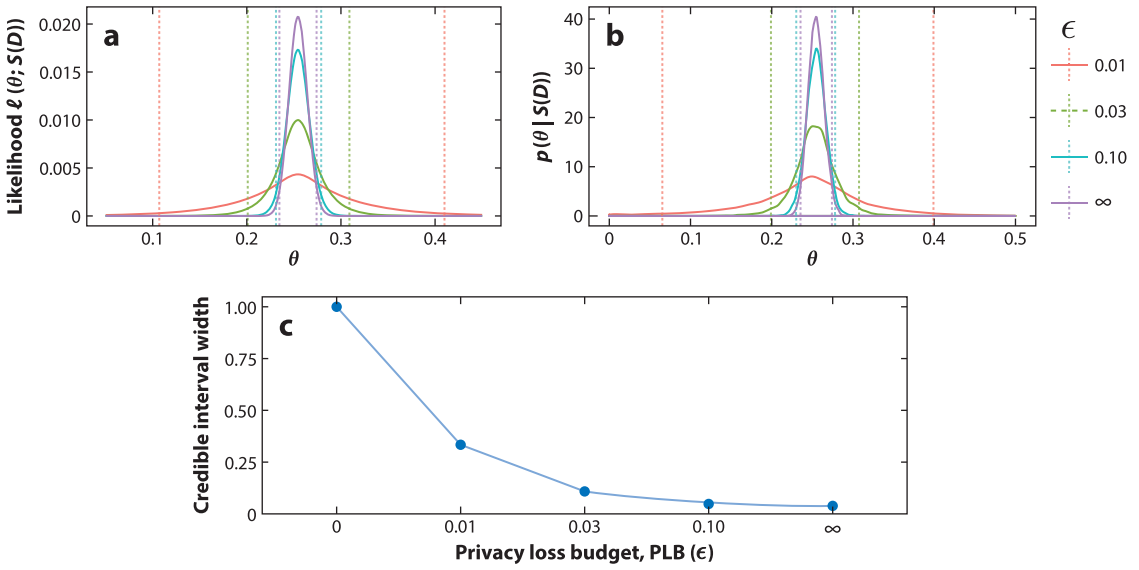

Figure 5

(*a*) Naive inference: probability of survival θ given sanitized statistic $S(D)$ = confidential statistic $T(D)$ under $T(D) \sim \text{Binomial}(n, \theta)$, and associated confidence intervals at different privacy loss budgets (PLBs) ϵ. (*b*) Adjusted inference: posterior distributions and associated credible intervals for $\theta|S(D)$ at different PLBs. (*c*) Risk-utility curve for private posterior inference on the *Game of Thrones* survival rate.

Most importantly, **Figure 5*c*** tells LV that posterior credible intervals for $\theta|S(D)$ under the Jeffreys prior $\theta \sim \text{Beta}(0.5, 0.5)$. Because LV properly accounted for errors due to privacy in his inference, his resulting credible interval increases in width as ϵ decreases while providing exact statistical coverage. This is not true in the case when we naively substitute $T(D)$ with $S(D)$, demonstrating once again the essential nature of statistically valid inference for sanitized results. Moreover, LV plotted this credible interval length on the y-axis with the PLB on the x-axis. This allows him to visualize the trade-off between privacy and utility and choose a PLB. LV sees that even the nonprivate result, $\epsilon = \infty$, has a small amount of uncertainty. By sacrificing a little data utility by making his credible interval wider for the number of dead in *GoT*, LV can help protect the citizens of Westeros from LF, regardless of whether it was the Dornish princess or someone else who completed the census of Westeros (see the sidebar titled The Morals of Our Fable). Huzzah!

THE MORALS OF OUR FABLE

- SDC measures can provide strong or weak privacy protections that scale differently with database sizes, and they may sometimes capture population-level effects that exist regardless of whose data contributed to the population-level inferences.
- All SDP methods affect downstream inference, even SDC methods that do not involve randomized noise.
- Although DP methods offer worst-case relative privacy guarantees, the posterior disclosure risks look different for different members of the database and different adversarial prior assumptions.
- DP necessitates adjusting downstream inferences for errors due to privacy, which requires analyzing the interaction of probability models with statistical measures.

6. DISCUSSION AND FUTURE DIRECTIONS

SDC and DP are schools of thought that frame the underlying problems of data privacy in different ways, as there are theoretical and empirical pros and cons to both approaches. Moreover, choosing to frame data privacy problems from one perspective or the other induces trade-offs that cannot always be quantitatively captured; these may be better solved by the legal and normative literatures on data privacy. In this section, we discuss the high-level differences between these approaches along a few key dimensions.

6.1. Comparing Statistical Disclosure Control and Differential Privacy: Quantifying Risk

SDC and DP rely on measures of privacy risk with different conceptual trade-offs, as discussed in Section 1.2. How can we tell which framework is better suited for a particular use case? For that, we need to think about the gaps left from either perspective. For SDC, the main question left unanswered is whether bounds on the DRMs allow for resilience against other kinds of inference attacks; for example, by using database reconstruction attacks on public Census data, the Census Bureau was able to identify real vulnerabilities in their previous SDC methodology (Garfinkel et al. 2019). For DP, the main question left unanswered is how to sociologically interpret PLBs, requiring us to reason about worst-case adversaries, database pairs, and disclosure scenarios. This shift in language can make it hard to express privacy concerns in terms of PLBs (Cummings et al. 2021). Making PLBs more interpretable, though, often requires further assumptions. For example, using the Bayesian formulation of ϵ-DP, we can make prior assumptions and calculate different posterior disclosure risks under such protections (McClure & Reiter 2012). Such measures offer more interpretability to practitioners at the expense of no longer providing worst-case guarantees.

Interpreting SDC and DP guarantees depends on many different considerations. How amenable is our output statistic space to privacy-preserving inference? How sociologically sensitive are the attributes about units we observe? How large is our database? What kinds of variability do we expect in the attribute responses? Sections 4 and 5 only begin to scratch the surface of answering these questions with respect to database size and query selection. Still, we demonstrate that answers to these questions change the social calculus of how we aim to quantify and limit privacy risks.

Future work could address an alternative approach, where the unit of analysis is neither a single database (as in SDC) nor an entire schema (as in DP), but instead a restricted set of database pairs within the schema. This approach, sitting somewhere between SDC and DP, could prove useful by making the DP-style worst-case-scenario analysis for DP less extreme while still providing more robustness to database reconstruction than SDC. Examples of early work in this area include Kifer & Machanavajjhala (2014), Song et al. (2017), and Seeman et al. (2022).

6.2. Comparing Statistical Disclosure Control and Differential Privacy: Quantifying Utility and Uncertainty

The role of data dependence is another distinguishing factor in comparing SDC and DP. Recall that SDC methods aim to formalize privacy as a property of a particular database $D \in \mathcal{D}$, whereas DP methods aim to formalize privacy as a property of a particular release process on a database schema $\mathcal{D}$. This change captures an important shortcoming of SDC methods. The way they are implemented could not be disclosed transparently without revealing probabilistic information about the individuals whose data were altered due to disclosure limitation. Still, many

optimal DP mechanisms rely on privacy-preserving errors in ways that depend heavily on the confidential data (e.g., Reimherr & Awan 2019, Asi & Duchi 2020), making $S(D) \mid D$ difficult to characterize in practice.

Should the distributions of randomized errors due to privacy depend on the confidential data? Even though the form of the mechanism can be transparently disclosed, the usefulness of this disclosure varies substantively for different mechanisms, which we explore in Section 4.3. From these examples, the discrete Laplace mechanism provides independent perturbations to collections of statistics; because the perturbation forms a location family with independent noise set by the PLB, the distance between the private statistic S and the nonprivate statistic $T(X)$ is independent of X. Aside from this relatively simple class of mechanisms, this property is not generally shared. Some primitive mechanisms, such as the exponential mechanism and its variants, do not easily allow for characterizations of errors due to privacy independent of the data. This is not to say we should not use mechanisms like these—it only means we should not ignore the tractability of valid downstream inference as a design consideration.

In particular, the ubiquitous use of postprocessing in DP methodology yields many different methods that meet certain optimality criteria, but for which the distribution of $S(D) \mid D$ is highly data dependent and sometimes computationally intractable. This is the case for the US Census TopDown algorithm, which sequentially postprocesses dependent count queries to conform to global public information and various internal self-consistency rules (Abowd et al. 2019). This suggests that both theoretically (Seeman et al. 2022) and empirically (Seeman et al. 2020), DP results should be released with and without postprocessing applied whenever possible.

6.3. Challenges in Schema Choice and Data Generation

SDP guarantees, regardless of whether using SDC or DP, depend heavily on the schema, $\mathcal{D}$. While SDP focuses on the form of the statistics we want to release, $\mathcal{S}$, the choice of $\mathcal{D}$ limits the possible values of $\mathcal{S}$ a priori. Moreover, from a system-level perspective, we tend to view $\mathcal{D}$ as a static entity, when in reality, schemas are dynamic and change over time. Schemas can grow to account for new unit attributes; for example, many databases containing protected health information are now updated to include information on COVID-19, such as vaccination status and testing history. Additionally, individual contributions to a database change over time, such as with streaming user data, which is an important consideration for databases regularly updated with event data, such as application logs from user behavior within different software applications. While there is some emerging work on this topic, we feel that neither SDC nor DP methodology has developed robust solutions to these problems yet. Hence, we see this as a budding area for future research opportunities.

Furthermore, SDP frameworks tend to view collected data as complete, full-information data, but rarely is this true in practice. Any social science data collection scheme could suffer from one of the many sources of total survey error (Groves et al. 2011), such as measurement error due to social desirability bias, errors due to missingness or other systematic nonresponse, or sampling procedures used to construct the database. We included these at the top of **Figure 1**, as most SDP analyses deal with human-level data. Because all information in SDP is typically taken at face value, the practical effects of accounting for ambiguity in this process are often lost.

Model-based SDC methods, like those discussed in Section 2.2, can account for some aspects of the data generating process, like survey sampling. However, incorporating similar ideas into DP is conceptually challenging, as the methodological details themselves also depend on the confidential data (Bun et al. 2020, Seeman & Brummet 2021). Resolving these differences is especially important for the needs of data curators at official statistical agencies like the US Census.

6.4. Privacy and Other Ethical Dimensions of Data Sciences

Even though we have focused on data privacy in a narrow, technical sense, privacy is a naturally interdisciplinary topic that involves philosophical, legal, and political scholarly traditions. The legal operationalization of SDP remains an open problem, as there is much debate as to how SDP approaches capture different legal statutes. Rogaway (2015) argues against any approach that a priori privileges one conception of privacy over another, as all SDP methods are inherently political in the way they allocate access to different data in different forms. If we argue one kind of political allocation is automatically better than another, we risk ignoring how defining the terms of that allocation influences our comparisons. Science and technology studies scholars refer to these as abstraction traps, which have been studied in algorithmic fairness (Selbst et al. 2019).

Additionally, SDP is but one of many research areas that attempts to imbue data analysis processes with sociologically desirable properties, such as interpretability (Carvalho et al. 2019) or fairness (Mitchell et al. 2021). Current research has pointed to limits in the ability to jointly satisfy DP guarantees and certain definitions of algorithmic fairness, both quantitatively (Cummings et al. 2019) and qualitatively (Green 2022).

6.5. Closing the Gap Between Theory and Practice

Here, we propose directions for open research that aims to resolve ideological tensions within SDC and DP research communities and direct future research toward addressing the needs of data subjects, curators, and users simultaneously.

SDP research, in its current state, is largely focused on establishing theoretical asymptotic results. Such results are clearly valuable, as they bound the sample complexity of SDP problems in collecting and releasing private statistics. However, when practitioners are deciding which method to use, we argue that such approaches fail to support those making such decisions, except for a very select few. Private companies that collect data at scale and use DP, like Google and Microsoft, have enough data to estimate the regimes in which asymptotic results offer useful characterizations of privacy risk and utility. But for small datasets, like many of those in the social and behavioral sciences, such techniques are infeasible. As a research community, we ought to enable everyone to use SDP, regardless of database size.

None of our critiques should detract from the theoretical value of this work, as it is an important step toward applicability. Instead, here we highlight open questions within SDP research that take into consideration practitioners' barriers to using SDP seriously. These research directions require substantive efforts from the SDP community (computer scientists, statisticians, and data users) to help close the widening gap between SDP theory and practice, most importantly along a few key dimensions:

1. Finite-sample utility guarantees: The close theoretical intersections between learning theory and privacy theory have motivated the sample complexity approach (i.e., analyzing DP mechanisms in terms of their asymptotic error guarantees as a function of database size and dimension) for particular SDP problems. As discussed above, this does not help practitioners easily identify the asymptotic regime in which these results apply. Future research ought to highlight tools that allow researchers working without data at scale to select optimal mechanisms for their use case.
2. Uncertainty quantification: Uncertainty quantification of various sources of errors is the foundation of statistical reasoning, and future SDP work needs to prioritize valid uncertainty quantification accounting for both errors in the data generating process (e.g., survey error) and those due to the privacy-preserving mechanism (e.g., Gaussian noise). For the design approach, this requires considering optimal inference in terms of total uncertainty, and

not just uncertainty about the nonprivate estimator. For the adjustment approach, this requires examining how SDP methods influence the bias and variability of statistics produced from sanitized results.

3. Optimization against operational intangibles: In theory, we tend to consider optimal mechanisms over a wide range of possible mechanisms that satisfy a particular privacy guarantee. In practice, though, only a subset of those mechanisms may be operationally feasible to meet data steward needs. For example, the US Census's requirements for releasing self-consistent microdata poses problems not only for optimality but also for consistent data utility across queries (Abowd et al. 2021). Future research should treat seriously these operational requirements, like the need for microdata or interpretable error distributions.
4. Computational barriers: The focus of the majority of DP mechanisms is in optimizing the trade-off between privacy and utility. However, computational issues are usually a third, neglected dimension of the problem, as mechanisms that are optimal from a privacy-utility perspective may be computationally prohibitive to implement. These problems arise deterministically with finite computing problems (Mironov 2012) as well as computing with randomized algorithms (Ganesh & Talwar 2020, Seeman et al. 2021). For example, instance-optimal mechanisms like the inverse sensitivity and K-norm gradient mechanisms require sampling from an intractable distribution, and failure to draw an exact sample using finite computing or finite MCMC approximation consumes additional PLB. Future research should explore trade-offs from this three-dimensional perspective instead of the two-dimensional perspective offered by privacy versus utility alone.
5. Extended trust models: Many models in privacy and security studies make different assumptions about trust, i.e., which parties have access to the confidential data and how. SDP tends to focus on one particular kind of trust model, the central model, in which the data curator is trusted to aggregate the data while respecting whatever privacy notion happens to be enforced. However, techniques from secure multiparty computation could be used to extend SDP methodology to offer more practical flexibility in trust modeling by relying on the distributed setting, where different physical machines are each tasked with different parts of the data processing and there are strict rules for how they communicate with one another (Karr 2010). There are many possible opportunities to synthesize studies of privacy-preserving secure multiparty computation and distributed analyses—that is, federated learning (e.g., Lindell & Pinkas 2009, Snoke et al. 2018a, Kairouz et al. 2021).

7. CONCLUSION

In conclusion, this review highlights and demonstrates the common methodological foundation of SDC and DP, and their associated quantitative and qualitative trade-offs required to investigate data privacy from either perspective. By focusing on the statistical viewpoint, SDP will produce and support the data sharing necessary for reproducible scientific discourse and democratic data governance. Whether using SDC or DP, or whether by design or adjustment, we all ought to remember that "different roads sometimes lead to the same castle" (Martin 2011, p. 95).

SUMMARY POINTS

1. Statistical disclosure control (SDC) and differential privacy (DP) methods are built upon common statistical foundations that make different but necessary compromises in conceptualizing privacy as properties of a particular database or as a schema.

2. Statistical data privacy (SDP) is inseparable from the study of data generating processes, as mechanism implementations introduce new privacy-preserving errors to be treated holistically alongside other error sources.
3. Both SDC and DP can suffer from model misspecification, and addressing this misspecification statistically can help improve our understanding of privacy and utility guarantees.
4. Future SDP research should address open statistical problems typically left unarticulated by theoretical SDP research, such as valid statistical inference, computational tractability, and compatibility with probability models, and their interplay.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

A.S. would like to acknowledge Vishesh Karwa and the late Stephen E. Fienberg for many early discussions and sharing of ideas on validity of privacy-preserving statistical inference; this article is in honor to them and to our joint work that was never completed. The authors are supported in part by National Science Foundation (NSF) Awards No. SES-1853209 and NCSES-BAA 49100421C0022 to The Pennsylvania State University.

LITERATURE CITED

Abowd J, Ashmead R, Cumings-Menon R, Garfinkel S, Kifer D, et al. 2021. An uncertainty principle is a price of privacy-preserving microdata. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, ed. M Ranzato, A Beygelzimer, Y Dauphin, PS Liang, J Wortman Vaughan. Red Hook, NY: Curran

Abowd J, Kifer D, Moran B, Ashmead R, Sexton W. 2019. *Census TopDown algorithm: differentially private data, incremental schemas, and consistency with public knowledge*. Work. Pap., US Census Bur., Washington, DC

Abowd JM. 2021. Third declaration of John M. Abowd. In *Fair Lines America Foundation, Inc. v. United States Department of Commerce and United States Bureau of the Census*, Civ. A. No. 1:21-cv-01361

Arnold C, Neunhoeffer M. 2020. Really useful synthetic data—a framework to evaluate the quality of differentially private synthetic data. arXiv:2004.07740 [stat.ML]

Asi H, Duchi JC. 2020. Instance-optimality in differential privacy via approximate inverse sensitivity mechanisms. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, ed. H Larochelle, M Ranzato, R Hadsell, MF Balcan, H Lin. Red Hook, NY: Curran

Avella-Medina M. 2021. Privacy-preserving parametric inference: a case for robust statistics. *J. Am. Stat. Assoc.* 116(534):969–83

Awan J, Kenney A, Reimherr M, Slavković A. 2019. Benefits and pitfalls of the exponential mechanism with applications to Hilbert spaces and functional PCA. *PMLR* 97:374–84

Awan J, Slavković A. 2018. Differentially private uniformly most powerful tests for binomial data. In *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, ed. H Wallach, H Larochelle, A Beygelzimer, F d'Alché-Buc, E Fox, R Garnett. Red Hook, NY: Curran

Awan J, Slavković A. 2020. Structure and sensitivity in differential privacy: Comparing *K*-norm mechanisms. *J. Am. Stat. Assoc.* 116:935–54

Beaumont MA. 2019. Approximate Bayesian computation. *Annu. Rev. Stat. Appl.* 6:379–403

Boucheron S, Lugosi G, Massart P. 2013. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford, UK: Oxford Univ. Press

Boulemtafes A, Derhab A, Challal Y. 2020. A review of privacy-preserving techniques for deep learning. *Neurocomputing* 384:21–45

Bousquet O, Boucheron S, Lugosi G. 2003. Introduction to statistical learning theory. In *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2–14, 2003, Tübingen, Germany, August 4–16, 2003, Revised Lectures*, ed. O Bousquet, U Luxburg, G Rätsch, pp. 169–207. New York: Springer

Bowen CM, Snoke J. 2019. Comparative study of differentially private synthetic data algorithms from the NIST PSCR Differential Privacy Synthetic Data Challenge. arXiv:1911.12704 [stat.AP]

boyd d, Sarathy J. 2022. Differential perspectives: epistemic disconnects surrounding the US Census Bureau's use of differential privacy. *Harv. Data Sci. Rev.* **https://doi.org/10.1162/99608f92.66882f0e**

Bun M, Drechsler J, Gaboardi M, McMillan A. 2020. Controlling privacy loss in survey sampling. arXiv:2007.12674 [stat.ME]

Bun M, Steinke T. 2016. Concentrated differential privacy: simplifications, extensions, and lower bounds. arXiv:1605.02065 [cs.CR]

Canonne CL, Kamath G, Steinke T. 2020. The discrete Gaussian for differential privacy. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, ed. H Larochelle, M Ranzato, R Hadsell, MF Balcan, H Lin, pp. 14106–17. Red Hook, NY: Curran

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. 2006. *Measurement Error in Nonlinear Models*. Boca Raton, FL: Chapman and Hall

Carvalho DV, Pereira EM, Cardoso JS. 2019. Machine learning interpretability: a survey on methods and metrics. *Electronics* 8(8):832

Chaudhuri K, Monteleoni C, Sarwate AD. 2011. Differentially private empirical risk minimization. *J. Mach. Learn. Res.* 12(3):1069–109

Chaudhuri K, Sarwate A, Sinha K. 2012. Near-optimal differentially private principal components. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, ed. F Pereira, CJ Burges, L Bottou, KQ Weinberger. Red Hook, NY: Curran

Cohen JE. 2012. What privacy is for. *Harv. Law Rev.* 126:1904–33

Cummings R, Gupta V, Kimpara D, Morgenstern J. 2019. On the compatibility of privacy and fairness. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, pp. 309–15. New York: ACM

Cummings R, Kaptchuk G, Redmiles EM. 2021. "I need a better description": an investigation into user expectations for differential privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, pp. 3037–52. New York: ACM

Dalenius T. 1977. Towards a methodology for statistical disclosure control. *Stat. Tidskr.* 15:429–44

Desfontaines D, Pejó B. 2022. SoK: differential privacies. arXiv:1906.01337 [cs.CR]

Dinur I, Nissim K. 2003. Revealing information while preserving privacy. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pp. 202–10. New York: ACM

Dobra A, Fienberg SE, Rinaldo A, Slavkovic A, Zhou Y. 2009. Algebraic statistics and contingency table problems: log-linear models, likelihood estimation, and disclosure limitation. In *Emerging Applications of Algebraic Geometry*, ed. M Putinar, S Sullivant, pp. 63–88. New York: Springer

Domingo-Ferrer J, Mateo-Sanz JM. 2002. Practical data-oriented microaggregation for statistical disclosure control. *IEEE Trans. Knowl. Data Eng.* 14(1):189–201

Domingo-Ferrer J, Sánchez D, Blanco-Justicia A. 2021. The limits of differential privacy (and its misuse in data release and machine learning). *Commun. ACM* 64(7):33–35

Domingo-Ferrer J, Torra V. 2003. Disclosure risk assessment in statistical microdata protection via advanced record linkage. *Stat. Comput.* 13(4):343–54

Dong J, Roth A, Su WJ. 2019. Gaussian differential privacy. arXiv:1905.02383 [cs.LG]

Drechsler J, Reiter JP. 2010. Sampling with synthesis: a new approach for releasing public use census microdata. *J. Am. Stat. Assoc.* 105(492):1347–57

Duchi JC, Jordan MI, Wainwright MJ. 2018. Minimax optimal procedures for locally private estimation. *J. Am. Stat. Assoc.* 113(521):182–201

Duncan GT, Pearson RW. 1991. Enhancing access to microdata while protecting confidentiality: prospects for the future. *Stat. Sci.* 6(3):219–32

Dwork C, Kenthapadi K, McSherry F, Mironov I, Naor M. 2006a. Our data, ourselves: privacy via distributed noise generation. In *Advances in Cryptology—EUROCRYPT 2006*, ed. S Vaudenay, pp. 486–503. New York: Springer

Dwork C, Lei J. 2009. Differential privacy and robust statistics. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing, STOC '09*, pp. 371–80. New York: ACM

Dwork C, McSherry F, Nissim K, Smith A. 2006b. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006*, ed. S Halevi, T Rabin, pp. 265–84. New York: Springer

Dwork C, Naor M. 2010. On the difficulties of disclosure prevention in statistical databases or the case for differential privacy. *J. Priv. Confid.* 2(1). **https://doi.org/10.29012/jpc.v2i1.585**

Dwork C, Roth A. 2014. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.* 9(3–4):211–407

Dwork C, Rothblum GN, Vadhan S. 2010. Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pp. 51–60. New York: IEEE

Dwork C, Smith A, Steinke T, Ullman J. 2017. Exposed! A survey of attacks on private data. *Annu. Rev. Stat. Appl.* 4:61–84

Evfimievski A, Gehrke J, Srikant R. 2003. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pp. 211–22. New York: ACM

Fearnhead P, Prangle D. 2012. Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *J. R. Stat. Soc. Ser. B* 74(3):419–74

Fienberg SE, Steele RJ. 1998. Disclosure limitation using perturbation and related methods for categorical data. *J. Off. Stat.* 14(4):485–502

Foulds J, Geumlek J, Welling M, Chaudhuri K. 2016. On the theory and practice of privacy-preserving Bayesian data analysis. arXiv:1603.07294 [cs.LG]

Franconi L, Polettini S. 2004. Individual risk estimation in μ-Argus: a review. In *Privacy in Statistical Databases CASC Project International Workshop, PSD 2004*, pp. 262–72. New York: Springer

Ganesh A, Talwar K. 2020. Faster differentially private samplers via Rényi divergence analysis of discretized Langevin MCMC. *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, ed. H Larochelle, M Ranzato, R Hadsell, MF Balcan, H Lin. Red Hook, NY: Curran

Garfinkel S, Abowd J, Martindale C. 2019. Understanding database reconstruction attacks on public data. *Commun. ACM* 62:46–53

Ghosh A, Roughgarden T, Sundararajan M. 2012. Universally utility-maximizing privacy mechanisms. *SIAM J. Comput.* 41(6):1673–93

Gong R. 2022. Exact inference with approximate computation for differentially private data via perturbations. arXiv:1909.12237 [stat.CO]

Green B. 2022. Escaping the impossibility of fairness: from formal to substantive algorithmic fairness. arXiv:2107.04642 [cs.CY]

Groves RM, Fowler FJ, Couper MP, Lepkowski JM, Singer E, Tourangeau R. 2011. *Survey Methodology*. New York: Wiley

Hardin JW, Hilbe JM. 2002. *Generalized Estimating Equations*. Boca Raton, FL: Chapman and Hall/CRC

Hardt M, Talwar K. 2010. On the geometry of differential privacy. In *STOC '10: Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, pp. 705–14. New York: ACM

Hundepool A, Domingo-Ferrer J, Franconi L, Giessing S, Nordholt ES, et al. 2012. *Statistical Disclosure Control*. New York: Wiley

Jordon J, Yoon J, van der Schaar M. 2019. Differentially private bagging: improved utility and cheaper privacy than subsample-and-aggregate. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, ed. H Wallach, H Larochelle, A Beygelzimer, F d'Alché-Buc, E Fox, R Garnett. Red Hook, NY: Curran

Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, et al. 2021. Advances and open problems in federated learning. *Found. Trends Mach. Learn.* 14(1–2):1–210

Karr AF. 2010. Secure statistical analysis of distributed databases, emphasizing what we don't know. *J. Priv. Confid.* 1:197–211

Karwa V, Krivitsky PN, Slavković AB. 2017. Sharing social network data: differentially private estimation of exponential family random-graph models. *J. R. Stat. Soc. Ser. C* 66(3):481–500

Karwa V, Slavković A. 2016. Inference using noisy degrees: differentially private β-model and synthetic graphs. *Ann. Stat.* 44(1):87–112

Karwa V, Vadhan S. 2017. Finite sample differentially private confidence intervals. arXiv:1711.03908 [cs.CR]

Kasiviswanathan SP, Smith A. 2014. On the 'semantics' of differential privacy: a Bayesian formulation. *J. Priv. Confid.* 6(1). **https://doi.org/10.29012/jpc.v6i1.634**

Kenny CT, Kuriwaki S, McCartan C, Rosenman E, Simko T, Imai K. 2021. The impact of the US Census Disclosure Avoidance System on redistricting and voting rights analysis. arXiv:2105.14197 [stat.AP]

Kifer D, Machanavajjhala A. 2011. No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, pp. 193–204. New York: ACM

Kifer D, Machanavajjhala A. 2014. Pufferfish: a framework for mathematical privacy definitions. *ACM Trans. Database Syst.* 39(1):3

Kifer D, Smith A, Thakurta A. 2012. Private convex empirical risk minimization and high-dimensional regression. *J. Mach. Learn. Res.* 23:25

Li N, Li T, Venkatasubramanian S. 2007. *t*-Closeness: privacy beyond *k*-anonymity and *l*-diversity. In *2007 IEEE 23rd International Conference on Data Engineering*, pp. 106–15. New York: IEEE

Li N, Qardaji W, Su D. 2012. On sampling, anonymization, and differential privacy or, *k*-anonymization meets differential privacy. In *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security*, pp. 32–33. New York: ACM

Lindell Y, Pinkas B. 2009. Secure Multiparty Computation for Privacy-Preserving Data Mining. *J. Priv. Confid.* 1(1):59–98

Liu C, Chakraborty S, Mittal P. 2016. *Dependence makes you vulnerable: differential privacy under dependent tuples*. Presented at Network and Distributed System Security Symposium, Feb. 21–24, San Diego, CA. **http://dx.doi.org/10.14722/ndss.2016.23279**

Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M. 2007. *L*-diversity: privacy beyond *k*-anonymity. *ACM Trans. Knowl. Discov. Data* 1(1):3

Martin GRR. 2011. *A Game of Thrones*. New York: Bantam

McClure D, Reiter JP. 2012. Differential privacy and statistical disclosure risk measures: an investigation with binary synthetic data. *Trans. Data Priv.* 5(3):535–52

McKenna L. 2019. *Disclosure avoidance techniques used for the 1960 through 2010 decennial censuses of population and housing public use microdata samples*. Work. Pap., US Census Bur., Washington, DC

McKenna R, Miklau G, Hay M, Machanavajjhala A. 2018. Optimizing error of high-dimensional statistical queries under differential privacy. *Proc. VLDB Endow.* 11(10):1206–19

McKenna R, Sheldon D, Miklau G. 2019. Graphical-model based estimation and inference for differential privacy. *PMLR* 97:4435–44

McSherry F, Talwar K. 2007. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pp. 94–103. New York: IEEE

McSherry FD. 2009. Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, pp. 19–30. New York: ACM

Minami K, Arai H, Sato I, Nakagawa H. 2016. Differential privacy without sensitivity. In *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, ed. D Lee, M Sugiyama, U Luxburg, I Guyon, R Garnett, pp. 956–64. Red Hook, NY: Curran

Mironov I. 2012. On significance of the least significant bits for differential privacy. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, pp. 650–61. New York: ACM

Mitchell S, Potash E, Barocas S, D'Amour A, Lum K. 2021. Algorithmic fairness: choices, assumptions, and definitions. *Annu. Rev. Stat. Appl.* 8:141–63

Nissim K, Raskhodnikova S, Smith A. 2007. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, pp. 75–84. New York: ACM

Ohm P. 2009. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Rev.* 57:1701–73

Raghunathan TE. 2021. Synthetic data. *Annu. Rev. Stat. Appl.* 8:129–40

Reimherr M, Awan J. 2019. KNG: The K-norm gradient mechanism. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, ed. H Wallach, H Larochelle, A Beygelzimer, F d'Alché-Buc, E Fox, R Garnett. Red Hook, NY: Curran

Rogaway P. 2015. *The moral character of cryptographic work.* Work. Pap. 1162, IACR Cryptol. ePrint Arch.

Seeman J, Brummet Q. 2021. *Posterior risk and utility from private synthetic weighted survey data*. Presented at World Meeting of the International Society for Bayesian Analysis (ISBA), virtual, June 28–July 2

Seeman J, Reimherr M, Slavković A. 2021. Exact privacy guarantees for Markov chain implementations of the exponential mechanism with artificial atoms. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, ed. M Ranzato, A Beygelzimer, Y Dauphin, PS Liang, J Wortman Vaughan. Red Hook, NY: Curran

Seeman J, Slavković A, Reimherr M. 2020. Private posterior inference consistent with public information: a case study in small area estimation from synthetic census data. In *Privacy in Statistical Databases: UNESCO Chair in Data Privacy*, *International Conference*, *PSD 2020*, ed. J Domingo-Ferrer, K Muralidhar, pp. 323–36. New York: Springer

Seeman J, Slavkovic A, Reimherr M. 2022. A formal privacy framework for partially private data. arXiv:2204.01102 [cs.CR]

Seeman J, Susser D. 2022. *Between privacy and utility*. Presented at Privacy Law Scholars Conference, Boston, June 2–3

Selbst AD, Boyd D, Friedler SA, Venkatasubramanian S, Vertesi J. 2019. Fairness and abstraction in sociotechnical systems. In *FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 59–68. New York: ACM

Skinner C. 2009. Statistical disclosure control for survey data. *Handb. Stat.* 29:381–96

Skinner CJ, Shlomo N. 2008. Assessing identification risk in survey microdata using log-linear models. *J. Am. Stat. Assoc.* 103:989–1001

Slavkovic A. 2013. Steve the matchmaker: the marriage of statistics and computer science in the world of data privacy. *CHANCE* 26(4):4–7

Slavkovic A, Molinari R. 2021. Perturbed *M*-estimation: a further investigation of robust statistics for differential privacy. In *Statistics in the Public Interest*, ed. AL Carriquiry, JM Tanur, WF Eddy, pp. 337–61. New York: Springer

Slavković AB. 2004. *Statistical disclosure limitation beyond the margins*. PhD Thesis, Carnegie Mellon Univ., Pittsburgh, PA

Slavković AB, Karwa V. 2019. *Statistical inference and privacy, part I*. Presented at Data Privacy: Foundations and Applications Boot Camp, Simons Inst., Berkeley, CA

Smith A. 2011. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, pp. 813–22. New York: ACM

Snoke J, Brick TR, Slavković A, Hunter MD. 2018a. Providing accurate models across private partitioned data: secure maximum likelihood estimation. *Ann. Appl. Stat.* 12(2):877–914

Snoke J, Raab GM, Nowok B, Dibben C, Slavkovic A. 2018b. General and specific utility measures for synthetic data. *J. R. Stat. Soc. Ser. A* 181(3):663–88

Snoke J, Slavković A. 2018. pMSE mechanism: differentially private synthetic data with maximal distributional similarity. In *Privacy in Statistical Databases: UNESCO Chair in Data Privacy*, *International Conference*, *PSD 2018*, ed. J Domingo-Ferrer, F Montes, pp. 138–59. New York: Springer

Song S, Chaudhuri K, Sarwate AD. 2013. Stochastic gradient descent with differentially private updates. In *2013 IEEE Global Conference on Signal and Information Processing*, pp. 245–48. New York: IEEE

Song S, Wang Y, Chaudhuri K. 2017. Pufferfish privacy mechanisms for correlated data. In *Proceedings of the 2017 ACM International Conference on Management of Data*, pp. 1291–306. New York: ACM

Stoller SD. 2011. Trust management in databases. In *Encyclopedia of Cryptography and Security*, ed. HCA van Tilborg, S Jajodia, pp. 1326–27. New York: Springer

Sweeney L. 2002. *k*-Anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowledge-Based Syst.* 10(5):557–70

Torkzadehmahani R, Kairouz P, Paten B. 2019. DP-CGAN: differentially private synthetic data and label generation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 98–104. New York: IEEE

Tsiatis AA. 2006. *Semiparametric Theory and Missing Data*. New York: Springer

Ullman J. 2021. *Statistical inference is not a privacy violation*. Differential Privacy Blog, June 3. **https://differentialprivacy.org/inference-is-not-a-privacy-violation/**

Vadhan S. 2017. The complexity of differential privacy. In *Tutorials on the Foundations of Cryptography*, ed. Y Lindell, pp. 347–450. New York: Springer

Vatsalan D, Christen P, Verykios VS. 2013. A taxonomy of privacy-preserving record linkage techniques. *Inform. Syst.* 38(6):946–69

Vershynin R. 2018. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge, UK: Cambridge Univ. Press

Vietri G, Tian G, Bun M, Steinke T, Wu S. 2020. New oracle-efficient algorithms for private synthetic data release. *PMLR* 119:9765–74

Vu D, Slavkovic A. 2009. Differential privacy for clinical trial data: preliminary evaluations. In *Proceedings of the 2009 IEEE International Conference on Data Mining Workshops, ICDMW '09*, pp. 138–43. New York: IEEE

Wang Y, Kifer D, Lee J, Karwa V. 2018. Statistical approximating distributions under differential privacy. *J. Priv. Confid.* 8(1). **https://doi.org/10.29012/jpc.666**

Wang YX, Fienberg S, Smola A. 2015. Privacy for free: posterior sampling and stochastic gradient Monte Carlo. *PMLR* 37:2493–502

Warner SL. 1965. Randomized response: a survey technique for eliminating evasive answer bias. *J. Am. Stat. Assoc.* 60(309):63–69

Wasserman L, Zhou S. 2010. A statistical framework for differential privacy. *J. Am. Stat. Assoc.* 105(489):375–89

Westin AF. 1968. Privacy and freedom. *Wash. Lee Law Rev.* 25(1):166

Willenborg L, De Waal T. 1996. *Statistical Disclosure Control in Practice*. New York: Springer

Winkler WE. 2004. Re-identification methods for masked microdata. In *Privacy in Statistical Databases*, ed. DF Josep, V Torra, pp. 216–30. New York: Springer