Prior-itizing Privacy: A Bayesian Approach to Setting the Privacy Budget in Differential Privacy

Zeki Kazan
Department of Statistical Science
Duke University
Durham, NC 27708
zekican.kazan@duke.edu

Jerome P. Reiter
Department of Statistical Science
Duke University
Durham, NC 27708
jreiter@duke.edu

Abstract

When releasing outputs from confidential data, agencies need to balance the analytical usefulness of the released data with the obligation to protect data subjects' confidentiality. For releases satisfying differential privacy, this balance is reflected by the privacy budget, ε. We provide a framework for setting ε based on its relationship with Bayesian posterior probabilities of disclosure. The agency responsible for the data release decides how much posterior risk it is willing to accept at various levels of prior risk, which implies a unique ε. Agencies can evaluate different risk profiles to determine one that leads to an acceptable trade-off in risk and utility.

1 Introduction

Differential privacy (DP) [11] is a gold standard definition of what it means for data curators, henceforth referred to as agencies, to protect individuals' confidentiality when releasing sensitive data. The confidentiality guarantee of DP is determined principally by a parameter typically referred to as the privacy budget ε. Smaller values of ε generally imply greater confidentiality protection at the cost of injecting more noise into the released data. Thus, agencies must choose ε to balance confidentiality protection with analytical usefulness. This balancing act has resulted in a wide range of values of ε in practice. For example, early advice in the field recommends ε of "0.01, 0.1, or in some cases, $\log(2)$ or $\log(3)$" [10], whereas recent large-scale implementations use $\varepsilon = 8.6$ in OnTheMap [31], $\varepsilon = 14$ in Apple's use of local DP for iOS 10.1.1 [44], and an equivalent of $\varepsilon = 17.14$ in the 2020 decennial census redistricting data release [1].

Some decision makers may find interpreting and selecting ε difficult [20, 24, 29, 36, 43]. For example, a recent study of practitioners using DP Creator [43] found that users wished for more explanation about how to select privacy parameters and better understanding of the effects of this choice. One potential path for providing guidance is to convert the task of selecting ε to setting bounds on the allowable probabilities of adversaries learning sensitive information from the released data, as in the classical disclosure limitation literature [8, 14, 41]. There is precedent for this approach: prior work in the literature suggests that "privacy semantics in terms of Bayesian inference can shed more light on a privacy definition than privacy semantics in terms of indistinguishable pairs of neighboring databases" [28] and prevailing advice for setting δ in (ε, δ)-DP originates from such Bayesian semantics [25].

We propose that agencies utilize relationships between DP and Bayesian semantics [2, 11, 25, 26, 27, 28, 29, 33] to select values of ε that accord with their desired confidentiality guarantees. The basic idea is as follows. First, the agency constructs a function that summarizes the maximum posterior probability of disclosure permitted for any prior probability of disclosure. For example, for a prior risk of 0.001, the agency may be comfortable with a ten-fold (or more) increase in the ratio of posterior risk to prior risk, whereas for a prior risk of 0.4, the agency may require the increase not exceed, say, 1.2. Second, for each prior risk value, the agency converts the posterior-to-prior ratio into the largest

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

ε that still ensures the ratio is satisfied. Third, the agency selects the smallest ε among these values, using that value for the data release. Importantly, the agency does not use the confidential data in these computations—they are theoretical and data free—so that they do not use up part of the overall privacy budget. Our main contributions include:

- We propose a framework for selecting ε under certain conditions (see Section 3) that applies to any DP mechanism, does not use additional privacy budget, and can account for disclosure risk from both an individual's inclusion and the sensitivity of values in the data.
- We enable agencies to tune the choice of ε to achieve their desired posterior-to-prior risk profile. This can avoid setting ε unnecessarily small if, for example, the agency tolerates larger posterior-to-prior ratios for certain prior risks. In turn, this can help agencies better manage trade-offs in disclosure risk and data utility.
- We give theoretic justification for the framework and derive closed-form solutions for the ε implied by several risk profiles. For more complex risk profiles, we also provide a general form for ε as a minimization problem.

To streamline the discussion, we focus on the release of discrete-valued statistics computed on discrete-valued data. Extension to continuous-valued statistics and data can be accomplished by replacing sums with integrals and PMFs with PDFs throughout the theorem statements and proofs.

2 Background and Motivation

We first describe some aspects of DP and Bayesian probabilities of disclosure relevant for our approach. We summarize all notation and definitions in Tables 2 and 3 in Appendix A.1.

2.1 Differential Privacy

Let $\mathbf{P}$ represent a population of individuals. The agency has a subset of $\mathbf{P}$, which we call $\mathbf{Y}$, comprising n individuals measured on d variables. For any individual i, let Y_i be the length-d vector of values corresponding to individual i, and let $I_i = 1$ when individual i is in $\mathbf{Y}$ and $I_i = 0$ otherwise. For all i such that $I_i = 1$, let $\mathbf{Y}_{-i}$ be the $(n-1) \times d$ matrix of values for the $n-1$ individuals in $\mathbf{Y}$ excluding individual i.[1] The agency wishes to release some function of the data, $T(\mathbf{Y})$. We assume $\mathbf{Y}$ and $T(\mathbf{Y})$ each have discrete support but may be many-dimensional. The agency turns to DP and will release $T^*(\mathbf{Y})$, a noisy version of $T(\mathbf{Y})$ under ε-DP.

Our work focuses on unbounded DP as defined in [9], which involves the removal or addition of one individual's information.[2] If a function $T^*(\mathbf{Y})$ with discrete support satisfies unbounded ε-DP, then for all i, all y in the support of Y_i, all $\mathbf{y}_{-i}$ in the support of $\mathbf{Y}_{-i}$, and all t^* in the support of $T^*(\mathbf{Y})$,

$$e^{-\varepsilon} \leq \frac{P[T^*(\mathbf{Y}) = t^* \mid Y_i = y, I_i = 1, \mathbf{Y}_{-i} = \mathbf{y}_{-i}]}{P[T^*(\mathbf{Y}_{-i}) = t^* \mid I_i = 0, \mathbf{Y}_{-i} = \mathbf{y}_{-i}]} \leq e^{\varepsilon}. \tag{1}$$

In settings where the statistic of interest is a count, a commonly used algorithm to satisfy DP is the geometric mechanism [18].

Definition 1 (Geometric Mechanism). *Let $T(\mathbf{Y}) \in \mathbb{Z}$ be a count statistic and suppose we wish to release a noisy count $T^*(\mathbf{Y}) \in \mathbb{Z}$ satisfying ε-DP. The geometric mechanism produces a count centered at $T(\mathbf{Y})$ with noise from a two-sided geometric distribution with parameter $e^{-\varepsilon}$. That is,*

$$P[T^*(\mathbf{Y}) = t^* \mid T(\mathbf{Y}) = t] = \frac{1 - e^{-\varepsilon}}{1 + e^{-\varepsilon}} e^{-\varepsilon|t^* - t|}, \qquad t^* \in \mathbb{Z}. \tag{2}$$

Under the geometric mechanism, the variance of $T^*(\mathbf{Y})$ is $2e^{-\varepsilon}/(1-e^{-\varepsilon})^2$. The variance increases as ε decreases, reflecting the potential loss in data usefulness when setting ε to a small value.

Of course, data usefulness is only one side of the coin. Agencies also need to assess the implications of the choice of ε for disclosure risks [7, 42, 43]. Prior works on setting ε focus on settings where (i)

[1] For all i such that $I_i = 0$ we let $\mathbf{Y}_{-i}$ be an $(n-1) \times d$ matrix of individuals not including individual i.

[2] This is in contrast to bounded DP [11], which involves the change of one individual's information.

the data have yet to be collected, and the goal is to simultaneously select ε and determine how much to compensate individuals for their loss in privacy [6, 15, 23, 30], (ii) the population is already public information, and the goal is to protect which subset of individuals is included in a release [29, 35], or (iii) data holders can utilize representative test data or the confidential data set itself [4, 17, 20, 35]. We focus on the common setting where data already have been collected, the population they are drawn from is not public information, and no data are available for tuning ε.

2.2 Bayesian Measures of Disclosure Risk

Consider an adversary who desires to learn about some particular individual i in $\mathbf{Y}$ using the release of $T^*(\mathbf{Y})$. We suppose that the adversary has a model, $\mathcal{M}$, for making predictions about the components of Y_i that they do not already know.[3] For example, the adversary could know some demographic information about individual i but not some other sensitive variable. The adversary might predict this sensitive variable from the demographic information using a model estimated with proprietary information or data from sources like administrative records. We assume that the release mechanism for $T^*(\mathbf{Y})$ is known to the adversary, that the DP release mechanism does not depend on $\mathcal{M}$, and that, under $\mathcal{M}$, the observations are independent but not necessarily identically distributed. These conditions are formalized in Section 3. For our ultimate purpose, i.e., helping agencies set ε, the exact form of the adversary's $\mathcal{M}$ is immaterial. In fact, as we shall discuss, we are not concerned whether the adversary's predictions from $\mathcal{M}$ are highly accurate or completely awful.

On a technical note, we make the distinction that the agency views $\mathbf{Y}$ and I_i as fixed quantities, since it knows which rows are in the collected data and what values are associated to each row. The adversary, however, views $\mathbf{Y}$ and I_i as random variables, and thus probabilistic statements about these quantities are well defined from the adversary's perspective. Notationally, we signify that a probabilistic statement is from the adversary's perspective via the subscript $\mathcal{M}$.

Let $\mathcal{S}$ be the subset of the support of Y_i that the agency considers a privacy violation. For example, if $d = 1$ and $\mathbf{Y}$ is income data, then $\mathcal{S}$ may be the set of possible incomes within $5{,}000$ or within 5% of the true income for individual i. Alternatively, if $d = 1$ and $\mathbf{Y}$ is binary, then $\mathcal{S}$ is a subset of $\{0, 1\}$. The selection of $\mathcal{S}$ must not depend on $\mathbf{P}$, as this might constitute a privacy violation.

The agency may be concerned about the risk that the adversary determines individual i is in $\mathbf{Y}$ or the risk that the adversary makes a disclosure for individual i; that is, $I_i = 1$ and $Y_i \in \mathcal{S}$, respectively. Assuming that the adversary's model puts nonzero probability mass on these events, we can express their relevant prior probabilities as follows.

$$P_{\mathcal{M}}[I_i = 1] = p_i, \qquad P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid I_i = 1] = q_i. \tag{3}$$

For fixed p_i and q_i, we can measure the risk of disclosure for individual i in a number of ways. Drawing from [33], one measure is the relative disclosure risk, $r_i(p_i, q_i, t^*)$. Writing the noisy statistic as T^* and suppressing the dependence on $\mathbf{Y}$ or $\mathbf{Y}_{-i}$, this is defined as follows.

Definition 2 (Relative Disclosure Risk). *For fixed data $\mathbf{Y}$, individual i, adversary's model $\mathcal{M}$, and released $T^* = t^*$, the relative disclosure risk is the posterior-to-prior risk ratio,*

$$r_i(p_i, q_i, t^*) = \frac{P_{\mathcal{M}}[Y_i \in \mathcal{S}, I_i = 1 \mid T^* = t^*]}{P_{\mathcal{M}}[Y_i \in \mathcal{S}, I_i = 1]}. \tag{4}$$

The relative risk can be decomposed into the posterior-to-prior ratio from inclusion (I_i) and the posterior-to-prior ratio from the values (Y_i). We have

$$r_i(p_i, q_i, t^*) = \frac{P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid I_i = 1, T^* = t^*]}{P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid I_i = 1]} \cdot \frac{P_{\mathcal{M}}[I_i = 1 \mid T^* = t^*]}{P_{\mathcal{M}}[I_i = 1]}. \tag{5}$$

The relative risk, however, does not tell the full story. As discussed in [22], the data holder also may care about absolute disclosure risks, $a_i(p_i, q_i, t^*)$.

Definition 3 (Absolute Disclosure Risk). *For fixed data $\mathbf{Y}$, individual i, adversary's model $\mathcal{M}$, and released $T^* = t^*$, the absolute disclosure risk is the posterior probability,*

$$a_i(p_i, q_i, t^*) = P_{\mathcal{M}}[Y_i \in \mathcal{S}, I_i = 1 \mid T^* = t^*]. \tag{6}$$

Since $r_i(p_i, q_i, t^*) = a_i(p_i, q_i, t^*)/(p_i q_i)$, we can convert between these risk measures.

[3] To be technically precise, we suppose the adversary assigns zero probability to predictions of Y_i with values of components that they know to be incorrect.

2.3 Risk Profiles

The quantities from Section 2.2 can inform the choice of ε. For example, DP implies that

$$r_i(p_i, q_i, t^*) \leq e^{2\varepsilon} \tag{7}$$

for all (p_i, q_i, t^*). [4] The inequality in (7) implies a naive strategy for setting ε: select a desired bound on the relative risk, r^*, and set $\varepsilon = \log(r^*)/2$. Practically, however, this strategy suffers from two drawbacks that could result in a smaller recommended ε than necessary. First, for any particular p_i and q_i, the bound in (7) need not be tight. In fact, this bound is actually quite loose across a wide range of values. Second, this strategy does not account for agencies willing to tolerate different relative risks for different prior probabilities. For example, if $p_i q_i = 0.25$, an agency may wish to limit the adversary's posterior to $a_i(p_i, q_i, t^*) \leq 2 \times 0.25 = 0.5$, but for $p_i q_i = 10^{-6}$, the same agency may find a limit of $a_i(p_i, q_i, t^*) \leq 2 \times 10^{-6}$ unnecessarily restrictive.

Rather than restricting to a constant bound, the agency can consider tolerable relative risks as a function of a hypothetical adversary's prior probabilities. We refer to this function as the agency's risk profile, denoted $r^*(p_i, q_i)$. Thus, the agency establishes a $r^*(p_i, q_i)$ so that, for all p_i, q_i, and t^*,

$$r_i(p_i, q_i, t^*) \leq r^*(p_i, q_i). \tag{8}$$

As the following results show, the requirement in (8) translates to a maximum value of ε.

3 Theoretical Results

In this section, we describe our main theoretical results. To devote more space to examples and applications, we describe supporting results in Appendix B.1 and provide all proofs in Appendix E.

We begin by formalizing the assumptions on the release mechanism for T^* and the adversary's model, $\mathcal{M}$. We emphasize that $P[T^*(\mathbf{Y}) = t^* \mid Y_i = y, \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 1]$ and $P[T^*(\mathbf{Y}_{-i}) = t^* \mid \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 0]$ are probabilities under the actual DP mechanism used for the release of $T^*(\mathbf{Y})$ when i is and is not included, respectively. We use these probabilities in two assumptions as follows.

Assumption 1. *For all y in the support of Y_i, $\mathbf{y}_{-i}$ in the support of $\mathbf{Y}_{-i}$, and t^* in the support of T^*,*

$$P_{\mathcal{M}}[T^*(\mathbf{Y}) = t^* \mid Y_i = y, \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 1] = P[T^*(\mathbf{Y}) = t^* \mid Y_i = y, \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 1]$$
$$P_{\mathcal{M}}[T^*(\mathbf{Y}_{-i}) = t^* \mid \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 0] = P[T^*(\mathbf{Y}_{-i}) = t^* \mid \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 0].$$

Assumption 1 implies that the mechanism for releasing T^* given $\mathbf{Y}$ is known to the adversary and that the adversary uses the actual release mechanism. An adversary who uses something other than the actual release mechanism to compute probabilities is unlikely in general to find more success than the adversary who does. Hence, Assumption 1 assures that the agency's computations are principled and cover a type of "worst-case" analysis.

Assumption 2. *For all y in the support of Y_i and $\mathbf{y}_{-i}$ in the support of $\mathbf{Y}_{-i}$,*

$$P_{\mathcal{M}}[\mathbf{Y}_{-i} = \mathbf{y}_{-i} \mid I_i = 1, Y_i = y] = P_{\mathcal{M}}[\mathbf{Y}_{-i} = \mathbf{y}_{-i} \mid I_i = 0]. \tag{9}$$

Assumption 2 implies that the adversary's beliefs about the distribution of $\mathbf{Y}_{-i}$ do not change whether individual i is included in the data or not, nor do they depend on individual i's confidential values. This assumption is similar to one explored in [28], who consider the case where "the presence/absence/record-value of each individual is independent of the presence/absence/record-values of other individuals," although we do not require that (9) holds for all i. Since DP is designed to capture the effect that a single atomic unit of data has on the output distribution, assuming this type of independence structure ensures one atomic unit is a single data point.[5] We believe these two assumptions are weaker than those in related work on choosing ε, which we discuss in Section 6.

Under Assumption 1 and Assumption 2, we can relate ε to the distribution of T^* unconditional on $\mathbf{Y}_{-i}$ via the following lemma. This result is similar, but not identical to, Theorem 6.1 in [28].

[4] This is proved in [19] under bounded DP in the case where $|\mathcal{S}| = 1$. We show in Corollary 1 in the appendix that, under our assumptions, it holds for larger sets and for unbounded DP (with ε replaced by 2ε).

[5] We thank a referee for highlighting this rationale for Assumption 2.

Lemma 1. *Under Assumption 1 and Assumption 2, if the release of $T^* = t^*$ satisfies ε-DP, then for any subset $\mathcal{S}$ of the domain of Y_i, we have*

$$e^{-\varepsilon} \leq \frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} \leq e^{\varepsilon}, \qquad e^{-2\varepsilon} \leq \frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1]} \leq e^{2\varepsilon}. \tag{10}$$

We emphasize that Lemma 1 and all subsequent results require Assumption 2. Without it, generalizable knowledge from the release can lead to arbitrarily large relative risk, as demonstrated by examples in [25, 27].

For a given function r^* selected by the data holder, we can determine the ε that should be used for the release. This is due to the following result relating the relative risk to ε.

Theorem 1. *Under Assumption 1 and Assumption 2, if the release of $T^* = t^*$ satisfies ε-DP, then*

$$r_i(p_i, q_i, t^*) \leq \frac{1}{q_i p_i + e^{-2\varepsilon}(1 - q_i)p_i + e^{-\varepsilon}(1 - p_i)}. \tag{11}$$

One can solve for $e^{-\varepsilon}$ in (11) to determine the recommended ε, which is given by Theorem 2.

Theorem 2. *For any individual i, fix the prior probabilities, p_i and q_i, and a desired bound on the relative disclosure risk, $r^*(p_i, q_i)$. Define $\varepsilon_i(p_i, q_i)$ to be the function*

$$\varepsilon_i(p_i, q_i) = \begin{cases} \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)}-p_iq_i\right)}-(1-p_i)}\right), & \textit{if } 0 < q_i < 1; \\ \log\left(\frac{1-p_i}{\frac{1}{r^*(p_i,1)}-p_i}\right), & \textit{if } q_i = 1. \end{cases} \tag{12}$$

Under the conditions of Theorem 1, any statistic $T^* = t^*$ released under ε-DP with $\varepsilon \leq \varepsilon_i(p_i, q_i)$ will satisfy $r_i(p_i, q_i, t^*) \leq r^*(p_i, q_i)$.

By Theorem 2, to achieve $r_i(p_i, q_i, t^*) \leq r^*(p_i, q_i)$ for all (p_i, q_i), the agency should set ε via the following minimization problem:

$$\varepsilon = \min_{p_i, q_i \in (0,1]} \varepsilon_i(p_i, q_i). \tag{13}$$

Closed forms for the ε resulting from a few specific risk profiles are included in the appendix.

4 Using Posterior-to-prior Risks for Setting ε

In this section, we illustrate how an agency can use the results of Section 3 to select ε. Notationally, for a given quantity x, we use $\tilde{x}$ to indicate the agency has fixed x to a particular constant.

Given $\mathcal{S}$, the agency must select a form for $r^*(p_i, q_i)$. A default choice, equivalent to the naive strategy using (7) discussed above, is to set the bound to a constant $\tilde{r} > 1$, i.e.,

$$r^*(p_i, q_i) = \tilde{r}. \tag{14}$$

As we prove in Theorem 3 in the appendix, the bound in (14) implies the agency should set $\varepsilon = \log(\tilde{r})/2$. While a constant bound on relative risk is simple, agencies that tolerate different risk profiles may be able to set ε to larger values, as we now illustrate.

Consider an agency that seeks to bound the relative risks for high prior probabilities and bound the absolute disclosure risk for low prior probabilities. For example, the agency may not want adversaries whose prior probabilities are low to use t^* to increase those probabilities beyond 0.10. Simultaneously, the agency may want to ensure adversaries with prior probabilities around, for example, 0.20 cannot use t^* to triple their posterior probability. Such an agency can specify a risk profile that requires either the relative risk be less than some $\tilde{r} > 1$ or the absolute risk be less than some $\tilde{a} < 1$, as we now illustrate via the following two examples.

The first example is a setting inspired by a case study in [13].

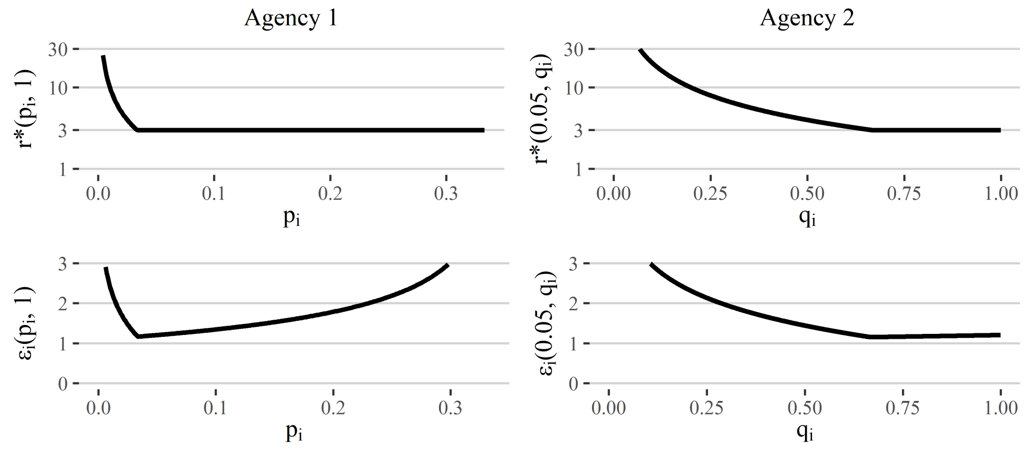

Figure 1: Each column corresponds to a particular hypothetical agency. The first row presents the agency's risk profile and the second row presents the profile's implied maximal allowable ε at each point on the curve. Agency 1's risk profile is given by (15) with $\tilde{a} = 0.1$, $\tilde{q} = 1$, and $\tilde{r} = 3$, while Agency 2's risk profile is given by (16) with $\tilde{a} = 0.1$, $\tilde{p} = 0.05$, and $\tilde{r} = 3$. For Agency 2, r^* and ε_i are very large for small q_i; large values are truncated for readability.

Agency	$\tilde{r}$	$\tilde{a}$	ε Recommendation	Noise Std. Dev.	Prob. Exact
A	1.5	0.25	0.51	2.74	25%
B	3	0.25	1.30	1.02	57%
C	6	0.25	2.04	0.59	77%

Table 1: The ε recommended by our framework for three risk profiles of the form in (15). Also included are the standard deviations of the noise distribution and the probabilities that the exact value is released, i.e., the mass at zero, for geometric mechanisms that satisfy the corresponding ε-DP.

Example 1. *A healthcare provider possesses a data set comprising demographic information about individuals diagnosed with a particular form of cancer in a region of interest. They plan to release the count of individuals diagnosed with this form of cancer in various demographic groups via the geometric mechanism, but are concerned this release, if insufficient noise is added, could reveal which individuals in the community have cancer. They wish to choose ε appropriately.*

In this example, the primary concern is with respect to inclusion in the data set. That is, for a given individual i, the adversary's prior probability p_i is the key quantity, whereas the q_i is not as important for any given $\mathcal{S}$. The agency can fix some $\tilde{q} \in (0, 1]$, for example $\tilde{q} = 1$ to focus on adversaries who already know individual i's demographic information. To simplify the analysis, we set the risk profile to ∞ for adversaries with $\tilde{q} \neq 1$, which indicates these adversaries are not considered. A reasonable risk profile for this agency then might be of the form, for some $\tilde{r} > 1$ and $\tilde{a} < 1$,[6]

$$r^*(p_i, q_i) = \begin{cases} \max\left\{\frac{\tilde{a}}{p_i\tilde{q}}, \tilde{r}\right\}, & \text{if } q_i = \tilde{q}; \\ \infty, & \text{if } q_i \neq \tilde{q}. \end{cases} = \begin{cases} \frac{\tilde{a}}{p_i\tilde{q}}, & \text{if } q_i = \tilde{q} \text{ and } 0 < p_i < \frac{\tilde{a}}{\tilde{q}\tilde{r}}; \\ \tilde{r}, & \text{if } q_i = \tilde{q} \text{ and } \frac{\tilde{a}}{\tilde{q}\tilde{r}} \leq p_i \leq 1; \\ \infty, & \text{if } q_i \neq \tilde{q}. \end{cases} \tag{15}$$

An example visualization of this risk function is presented in the first column of Figure 1. The upper plot displays the risk profile as a function of p_i when $q_i = \tilde{q}$, and the lower plot displays the maximal ε for which the relative risk bound holds for each p_i. We can derive a closed form for ε under this risk profile; see Table 4 in Appendix A.2 for a summary and Theorem 5 in Appendix B.1 for details.

Suppose the agency is generally willing to accept a maximum absolute disclosure risk of $\tilde{a} = 0.25$ for adversaries with small prior probabilities. Table 1 presents the maximal ε which satisfies the desired risk profile for three such agencies. Agency A is risk averse, agency C is utility seeking, and agency

[6] The second equality in (15) assumes $\tilde{q} > \tilde{a}/\tilde{r}$. If $\tilde{q} \leq \tilde{a}/\tilde{r}$, then $r^*(p_i, \tilde{q}) = \tilde{a}/(p_i\tilde{q})$ for all $0 < p_i \leq 1$.

B sits in between in terms of risk and utility. For adversaries with high prior probabilities, agencies A, B, and C bound the relative disclosure risk at $\tilde{r} = 1.5$, $\tilde{r} = 3$, and $\tilde{r} = 6$, respectively. Figure 3 in Appendix C.1 plots the forms of each agency's risk profile. To provide intuition on the amount of noise implied by these ε, in Table 1 we display the standard deviation of the noise distribution for each statistic under the geometric mechanism and the probability that the geometric mechanism will output the exact value of each statistic. The risk averse agency is recommended an ε that results in a release with a high standard deviation and fairly low probability of releasing the exact value of the statistic. The utility seeking agency is recommended an ε that results in a release with a fairly low standard deviation and high probability of releasing the exact value of the statistic.

The recommendations from these risk profiles, which are tailored to the setting and agency preferences, are higher than the recommendations from a corresponding simple risk profile of $r^*(p_i, q_i) = \tilde{r}$ for all prior probabilities. This simple profile yields $\varepsilon \approx 0.20$, $\varepsilon \approx 0.55$, and $\varepsilon \approx 0.90$ for $\tilde{r} = 1.5$, $\tilde{r} = 3$, and $\tilde{r} = 6$, respectively, less than half the corresponding ε values in Table 1.

We now consider a second example, inspired by Example 16 in [45], that alters Example 1.

Example 2. *A survey is performed on a sample of individuals in the region of interest. 5% of the region is surveyed, and respondents are asked whether they have this particular form of cancer along with a series of demographic questions. The agency plans to release the counts of surveyed individuals who have and do not have the cancer in various demographic groups via the geometric mechanism, but is concerned this release, if insufficient noise is added, could reveal which individuals in the community have cancer. The agency wishes to choose ε appropriately.*

In this example, the primary concern is with respect to the values in the data set. For ease of notation, let Y_i be a d-vector of binary values, and let the first element, Y_{1i}, be an indicator for whether individual i has this form of cancer. Set $\mathcal{S}$ to be the subset of the support of Y_i for which individual i has the cancer, i.e., $\mathcal{S} = \{y \in \{0,1\}^d : y_1 = 1\}$. For individual i, the adversary's q_i is the key quantity, and their p_i is not of interest. The agency can fix some $\tilde{p} \in (0, 1]$; for example, if the agency is most concerned about adversaries whose only prior knowledge is that individual i is in the population, but not whether they were surveyed, they might set $\tilde{p} = 0.05$. Alternatively, they might set $\tilde{p} = 1$ to imply an adversary that knows a priori that individual i is included in $\mathbf{Y}$. A reasonable risk profile for this agency might be of the form, for some $\tilde{a} < 1$ and $\tilde{r} > 1$, [7]

$$r^*(p_i, q_i) = \begin{cases} \max\left\{\frac{\tilde{a}}{\tilde{p}q_i}, \tilde{r}\right\}, & \text{if } p_i = \tilde{p}; \\ \infty, & \text{if } p_i \neq \tilde{p}. \end{cases} = \begin{cases} \frac{\tilde{a}}{\tilde{p}q_i}, & \text{if } p_i = \tilde{p} \text{ and } 0 < q_i < \frac{\tilde{a}}{\tilde{p}\tilde{r}}; \\ \tilde{r}, & \text{if } p_i = \tilde{p} \text{ and } \frac{\tilde{a}}{\tilde{p}\tilde{r}} \leq q_i \leq 1; \\ \infty, & \text{if } p_i \neq \tilde{p}. \end{cases} \tag{16}$$

An example visualization of this risk function is presented in the second column of Figure 1. The upper plot displays the risk profile as a function of q_i when $p_i = \tilde{p}$, and the lower plot displays the maximal ε for which the relative risk bound holds for each p_i. The agency can select the smallest ε on this curve for their release. We can derive a closed form the minimum; see Table 4 in Appendix A.2 for a summary and Theorem 4 in Appendix B.1 for details. Notably, it can be shown that the ε selected under these two profiles is bounded below by $\log(\tilde{r})/2$ and may be much larger. We provide further exploration and visualizations for this example in Appendix C.2.

As demonstrated by these examples, the gap between the baseline strategy of (14) and our recommendation can be large. To further illustrate, suppose an agency has a simple, "point" risk profile, where $r^*(0.5, 1) = r'$ for $r' < 2$ (and ∞ elsewhere). The baseline recommends $\varepsilon = \log(r')/2 < 0.35$, while the recommendation from Theorem 2 can be shown to have the form $\varepsilon = \log(r'/(2 - r'))$, which diverges as $r' \to 2$. Thus, the gap between the recommendations can be arbitrarily large. While this likely does not represent a realistic disclosure risk profile of a real-world agency, it is instructive. In general, the baseline's recommendation is smaller because it is enforcing a relative risk of r' for adversaries with small priors. If $r' < 2$, then an adversary with prior probability $p_i q_i = 0.001$ is restricted to a posterior probability at most 0.002, leading to a small ε recommendation. If adversaries with small priors are ignored or allowed to have large relative risks, the recommendation from our method will outperform the baseline, and possibly by a substantial amount.

A particular agency's risk profile may not be characterized by one of the forms described above. For example, an agency may be equally concerned about p_i and q_i, rather than focusing on one; see

[7] The second equality in (16) assumes $\tilde{p} > \tilde{a}/\tilde{r}$. If $\tilde{p} \leq \tilde{a}/\tilde{r}$, then $r^*(\tilde{p}, q_i) = \tilde{a}/(\tilde{p}q_i)$ for all $0 < q_i \leq 1$.

Appendix C.3 for an example of this setting. Or, as argued in [32], in some settings, agencies might be most concerned about the difference between the posterior and prior probabilities (rather than their ratio).[8] In such settings, it is straightforward to write down the corresponding risk function, $r^*(p_i, q_i)$, and the optimal ε can be determined by numerically solving the minimization problem in (13).[9]

Regardless of the agency's desiderata for a risk profile, we recommend that they keep the following in mind when setting its functional form. First, for any region where $r^*(p_i, q_i) > 1/(p_i q_i)$, the risk profile generates a bound on the posterior probability that exceeds 1. Of course, the posterior probabilities themselves cannot exceed 1; thus, in these regions, the risk profile effectively does not bound the posterior risk. For example, an agency that sets $r^*(p_i, 1) = 3$ in the region where $p_i \geq 1/3$ (as in the left column of Figure 1) implicitly is willing to accept an unbounded ε for prior probabilities $p_i \geq 1/3$. Second, when bounding the absolute disclosure risk below some $\tilde{a}$ in some region of (p_i, q_i), the agency should require $p_i q_i < \tilde{a}$ in that region. When $p_i q_i = \tilde{a}$, the recommended $\varepsilon = 0$ since the data holder requires T^* to offer no information about Y_i. This also suggests that an agency bounding absolute disclosure risk in a region of (p_i, q_i) that sets $\tilde{a}$ close to some value of $p_i q_i$ in the region is willing to accept only small ε values.

5 Managing the Trade-off in Privacy and Utility

Agencies can use the framework to manage trade-offs in disclosure risk and data utility. In particular, the agency can evaluate potential impacts of the DP algorithm using data quality metrics under different risk profiles, choosing an ε that leads to a satisfactory trade-off. We demonstrate this in the below example, using data on infant mortality in Durham county, N.C. We note that these data may not require the use of DP in reality, but they do allow us to illustrate the process with public data.

Example 3. *Suppose that an agency in Durham county, N.C., wishes to release a differentially private version of the number of infant deaths the county experienced in the year 2020. The agency plans to use the geometric mechanism to release this value. Of particular interest is whether the county infant death rate meets the U. S. Department of Health and Human Services' Healthy People 2020 target of* 6.0 *or fewer deaths per* 1,000 *live births [21]. The agency wishes to minimize the probability that they release a noisy count that changes whether or not the* 6.0 *target is met; their goal is to ensure this probability is below 10%. It is public information that there were* 4,012 *live births in Durham county in 2020 [38].*

The primary concern in Example 3 is with respect to inclusion in the data. Thus, the agency focuses on adversaries with $q_i = 1$ and varying p_i. The agency is generally willing to incur large relative risks if p_i is small and small relative risks if p_i is large. Furthermore, due to possible policy ramifications of appearing not to meet the target rate, the agency is open to accepting greater privacy risks for greater accuracy in the released count, within reason as determined by agency decision makers.

The agency's privacy experts determine that they can characterize the agency's risk profiles reasonably well using the following general class of risk profiles.

$$r^*(p_i, q_i) = \begin{cases} \max\left\{\frac{\tilde{a}}{p_i}, \tilde{r}\right\}, & \text{if } q_i = 1; \\ \infty, & \text{if } q_i \neq 1. \end{cases} \tag{17}$$

To allow for consideration of the effects on accuracy of different specifications of (17), the agency considers combinations of $\tilde{r} \in \{1.2, 2, 5\}$ and $\tilde{a} \in \{0.1, 0.25, 0.5\}$. They also examine $\tilde{a} = 0$, which corresponds to a constant risk profile $r^*(p_i, q_i) = \tilde{r}$. The agency could consider other combinations of parameters or risk profiles classes should these not lead to a satisfactory trade-off in risk and utility.

Figure 2 summarizes an analysis of the risk/utility trade-offs. For all $\tilde{r}$, the recommended ε is larger when $\tilde{a} > 0$ than under the corresponding constant risk profile. Similar increases in ε also are evident for the risk profiles of Section 4; see Appendix C. Naturally, increases in $\tilde{r}$ and $\tilde{a}$ decrease the RMSEs[10] of the noisy counts due to larger ε. The switching probabilities in the top panel do not use

[8] Requiring the difference between the posterior and prior probabilities to be bounded by some $\tilde{b} \in (0, 1)$ is equivalent to requiring the relative risk be bounded by $(p_i q_i + \tilde{b})/(p_i q_i)$. See Appendix B.2 for details.

[9] We provide `R` code at `https://github.com/zekicankazan/choosing_dp_epsilon` to determine the recommended ε for a provided risk profile. This code was used to determine closed forms for ε for the additional risk profiles described in Appendix B.2. The example in Appendix C.3 demonstrates the use of this code.

[10] Since T^* is unbiased for the true count, the RMSE is the standard deviation of the noise distribution.

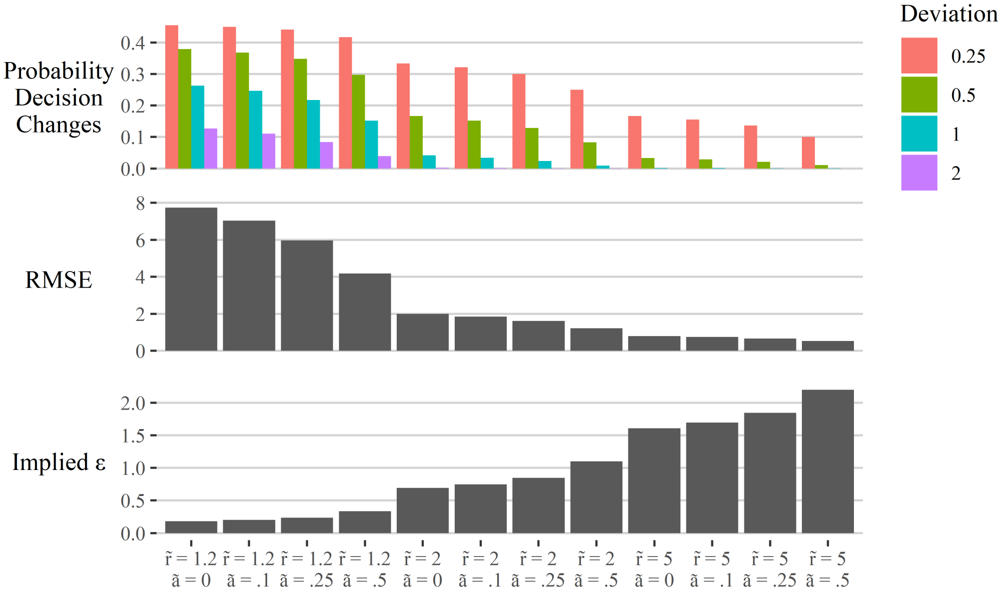

Figure 2: The top panel presents the probability that the differentially private algorithm switches whether Durham county's released rate is above or below the 6.0 target—e.g., the added noise makes the released rate 6.5 but the actual rate is 5.5. Each color represents a different hypothetical absolute difference between the true and target rate. The middle panel presents RMSEs of the noisy count of infant deaths. The bottom panel presents the implied ε. Each bar corresponds to a different risk profile of the form in (17).

the true count of infant deaths; rather, they consider hypothetical deviations of 0.25, 0.5, 1, and 2 in the true rate from the target rate of 6.0. In this way, the agency avoids additional privacy leakage. The probability differs substantially for each of the four deviations, with the deviation of 0.25 leading to the largest probability of a switch. To ensure the probability is below 10% for the deviation of 0.25, the agency would need to use the most aggressive of the risk profiles, which allows absolute disclosure risk of $\tilde{a} = 0.5$ or relative disclosure risk of $\tilde{r} = 5$. This leads to $\varepsilon \approx 2.20$ with RMSE ≈ 0.53. If they are willing to allow the probability to exceed 10% or to place priority on deviations exceeding 0.25, the agency could select a less aggressive risk profile that yields a smaller ε.

In actuality, there were 6.2 infant deaths per $1{,}000$ live births in Durham County in 2020 [37]. As the truth is close to the 6.0 target, the agency would need to use a large ε to achieve their goal.

6 Relationship to Prior Work

The most similar work we are aware of is Lee and Clifton [29]'s "How Much is Enough? Choosing ε for Differential Privacy," which also uses a Bayesian approach to set ε. The authors focus on settings where the population is public information and the adversary's goal is to determine which subset of individuals in the population was used for a differentially private release of a statistic. Their method for selecting ε is tailored to the setting where only disclosure of an individual's inclusion in the data is of concern and the values of the entire population can be used to inform the choice of ε. [29] focuses on the setting where the Laplace mechanism is used and the population size is known; their proposal was extended to any DP mechanism by [24, 39] and settings where the population size is unknown by [34]. In Appendix D.1, we compare the recommended ε from the proposal in [29] and our framework in several examples. We find that the approach of [29], for particular statistics and populations, can recommend a larger ε than the one resulting from our approach. However, our framework may be used in settings where the values of the entire population are not public and settings where the privacy of the values in the release is of primary concern.

Another series of related work involves using Bayesian semantics to communicate the meaning of ε to potential study participants. For an overview of recent work in this area, see [16], [36], and citations therein. We highlight an example in [45] (corrected in [46]), which considers an individual deciding whether or not to participate in a survey for which results will be released via DP with a particular ε. The authors suggest that the individual consider a bound on a posterior-to-posterior ratio similar to the bound in (11) to make an informed decision about whether to participate in the survey. We describe the advice of [45] using our notation in Appendix D.2. While their and our bounds have similar expressions, the goal in [45] differs from ours. They use the bound to characterize the individual's disclosure risks for a fixed ε, whereas we establish the agency's disclosure risk profile in order to set ε.

There are a number of other works that examine Bayesian semantics of ε-DP, which we now briefly summarize. In their seminal work proposing DP, [11] showed that, under some conditions, bounded DP is equivalent to a bound on the relative risk when all but one data point is fixed. [25] showed that bounded DP implies a bound on the total variation distance between the posterior distribution with all data points included and the posterior distribution with one data point removed. [28] showed that both bounded and unbounded DP are equivalent to a bound on the posterior-to-prior odds ratio exactly when the presence/absence/record-value of each individual is independent of that of other individuals. Finally, [27] showed that both bounded and unbounded DP imply bounds on the ratio of the posterior distribution with all data points included to the posterior with one observation replaced with a draw from the posterior with their observation removed. These works do not discuss how these semantic characterizations of DP can be used to select ε.

7 Commentary

In this article, we propose a framework for selecting ε for DP by establishing disclosure risk profiles. Essentially, we provide a method for agencies to trade the problem of selecting ε for a release for the problem of specifying their desired disclosure risk profile. This process involves focusing on particular classes of adversaries the agency is most concerned about—represented by the set $\mathcal{S}$—and tuning ε to ensure the risk from these adversaries is sufficiently low. We emphasize that, once applied, DP will protect against all attacks with the guarantee of DP, not just the attack used to tune ε.

We primarily focus on the risk from one class of adversary attacks on one individual, assuming all individuals are exchangeable. If agencies assign different risk profiles for different classes of adversaries, they could repeat the analysis for each class For example, the agency might assign different risk profiles to adversaries targeting different characteristics, e.g., they may consider whether an individual has a disease more sensitive than their age. Similarly, if an agency does not treat individuals in the data as exchangeable—for example, if an agency seeks to ensure lesser disclosure risks for groups with some characteristics than for not-necessarily-disjoint groups without those characteristics —the agency could repeat this analysis for each group. In both of these settings, the agency has a decision problem on its hands. As with designing DP solutions in general, the agency must prioritize some risks over others, e.g., using decision-theoretic criteria. This is an important topic for future work and raises ethical questions regarding how assigning different risk profiles for individuals with different characteristics could affect data equity.

One avenue for future work involves incorporating a version of this framework into differentially private data analysis tools. Recently developed interfaces for DP data releases such as [24] and [35] use the framework of [29] to guide users in setting ε. Our framework could be similarly incorporated to provide guidance in settings that do not satisfy the assumption of [29] that the values of the population are public. Another avenue for future work involves examining how an agency can set the risk profile appropriate for a particular release. We envision agencies could do so analogously to the elicitation of utility functions in decision theory, e.g., by considering a series of bets [40]. Other future extensions involve examining whether similar results follow under weaker assumptions. This includes settings with multiple differentially private releases via existing results that relate the relative risk to DP composition theorems (e.g., S5 of [26]). This also includes extension to variants of DP—such as zero-concentrated DP [5] or approximate DP [12]—by deriving Bayesian semantics analogous to Theorem 1 or leveraging existing results [25, 27]. Additionally, it may be possible to extend the results in this article from posterior-to-prior risks to the sorts of posterior-to-posterior risks discussed in Section 6.

Acknowledgments and Disclosure of Funding

This research was supported by NSF grant SES-2217456. We would like to thank the anonymous reviewers for their helpful feedback as well as Salil Vadhan and Jörg Drechsler for insightful discussion about a preliminary version of this work.

References

[1] John M Abowd, Robert Ashmead, Ryan Cumings-Menon, Simson Garfinkel, Micah Heineck, Christine Heiss, Robert Johns, Daniel Kifer, Philip Leclerc, Ashwin Machanavajjhala, et al. The 2020 census disclosure avoidance system topdown algorithm. *arXiv preprint arXiv:2204.08986*, 2022.

[2] John M Abowd and Lars Vilhuber. How protective are synthetic data? In *International Conference on Privacy in Statistical Databases*, pages 239–246. Springer, 2008.

[3] Borja Balle, Gilles Barthe, and Marco Gaboardi. Privacy amplification by subsampling: Tight analyses via couplings and divergences. *Advances in Neural Information Processing Systems*, 31, 2018.

[4] Mihai Budiu, Pratiksha Thaker, Parikshit Gopalan, Udi Wieder, and Matei Zaharia. Overlook: Differentially private exploratory visualization for big data. *Journal of Privacy and Confidentiality*, 12(1), 2022.

[5] Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, pages 635–658. Springer, 2016.

[6] Pranav Dandekar, Nadia Fawaz, and Stratis Ioannidis. Privacy auctions for recommender systems. *ACM Transactions on Economics and Computation (TEAC)*, 2(3):1–22, 2014.

[7] Jörg Drechsler. Differential privacy for government agencies—are we there yet? *Journal of the American Statistical Association*, 118(541):761–773, 2023.

[8] George T Duncan and Diane Lambert. Disclosure-limited data dissemination. *Journal of the American Statistical Association*, 81(393):10–18, 1986.

[9] Cynthia Dwork. Differential privacy. In *International colloquium on automata, languages, and programming*, pages 1–12. Springer, 2006.

[10] Cynthia Dwork. A firm foundation for private data analysis. *Communications of the ACM*, 54(1):86–95, 2011.

[11] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference*, pages 265–284. Springer, 2006.

[12] Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. Boosting and differential privacy. In *2010 IEEE 51st annual symposium on foundations of computer science*, pages 51–60. IEEE, 2010.

[13] Amalie Dyda, Michael Purcell, Stephanie Curtis, Emma Field, Priyanka Pillai, Kieran Ricardo, Haotian Weng, Jessica C Moore, Michael Hewett, Graham Williams, et al. Differential privacy for public health data: An innovative tool to optimize information sharing while protecting data confidentiality. *Patterns*, 2(12):100366, 2021.

[14] Stephen E Fienberg and Ashish P Sanil. A Bayesian approach to data disclosure: Optimal intruder behavior for continuous data. *Journal of Official Statistics*, 13(1):75, 1997.

[15] Lisa K Fleischer and Yu-Han Lyu. Approximately optimal auctions for selling privacy when costs are correlated with data. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 568–585, 2012.

[16] Daniel Franzen, Saskia Nuñez von Voigt, Peter Sörries, Florian Tschorsch, and Claudia Müller-Birn. Am i private and if so, how many? communicating privacy guarantees of differential privacy with risk communication formats. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, pages 1125–1139, 2022.

[17] Marco Gaboardi, James Honaker, Gary King, Jack Murtagh, Kobbi Nissim, Jonathan Ullman, and Salil Vadhan. Psi ({\Psi}): a private data sharing interface. *arXiv preprint arXiv:1609.04340*, 2016.

[18] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing*, 41(6):1673–1693, 2012.

[19] Ruobin Gong and Xiao-Li Meng. Congenial differential privacy under mandated disclosure. In *Proceedings of the 2020 ACM-IMS on Foundations of Data Science Conference*, pages 59–70, 2020.

[20] Michael Hay, Ashwin Machanavajjhala, Gerome Miklau, Yan Chen, Dan Zhang, and George Bissias. Exploring privacy-accuracy tradeoffs using dpcomp. In *Proceedings of the 2016 International Conference on Management of Data*, pages 2101–2104, 2016.

[21] HHS. Healthy people 2020. Office of Disease Prevention and Health Promotion, U.S. Department of Health and Human Services, 2020. Accessed on 22 April 2024.

[22] V Joseph Hotz, Christopher R Bollinger, Tatiana Komarova, Charles F Manski, Robert A Moffitt, Denis Nekipelov, Aaron Sojourner, and Bruce D Spencer. Balancing data privacy and usability in the federal statistical system. *Proceedings of the National Academy of Sciences*, 119(31), 2022.

[23] Justin Hsu, Marco Gaboardi, Andreas Haeberlen, Sanjeev Khanna, Arjun Narayan, Benjamin C Pierce, and Aaron Roth. Differential privacy: An economic method for choosing epsilon. In *2014 IEEE 27th Computer Security Foundations Symposium*, pages 398–410. IEEE, 2014.

[24] Mark F St John, Grit Denker, Peeter Laud, Karsten Martiny, Alisa Pankova, and Dusko Pavlovic. Decision support for sharing data using differential privacy. In *2021 IEEE Symposium on Visualization for Cyber Security (VizSec)*, pages 26–35. IEEE, 2021.

[25] Shiva P Kasiviswanathan and Adam Smith. On the'semantics' of differential privacy: A Bayesian formulation. *Journal of Privacy and Confidentiality*, 6(1), 2014.

[26] Zeki Kazan and Jerome Reiter. Assessing statistical disclosure risk for differentially private, hierarchical count data, with application to the 2020 us decennial census. *Statistica Sinica*, 34:to appear, 2024.

[27] Daniel Kifer, John M Abowd, Robert Ashmead, Ryan Cumings-Menon, Philip Leclerc, Ashwin Machanavajjhala, William Sexton, and Pavel Zhuravlev. Bayesian and frequentist semantics for common variations of differential privacy: Applications to the 2020 census. *arXiv preprint arXiv:2209.03310*, 2022.

[28] Daniel Kifer and Ashwin Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems (TODS)*, 39(1):1–36, 2014.

[29] Jaewoo Lee and Chris Clifton. How much is enough? choosing ε for differential privacy. In *International Conference on Information Security*, pages 325–340. Springer, 2011.

[30] Chao Li, Daniel Yang Li, Gerome Miklau, and Dan Suciu. A theory of pricing private data. *ACM Transactions on Database Systems (TODS)*, 39(4):1–28, 2014.

[31] Ashwin Machanavajjhala, Daniel Kifer, John Abowd, Johannes Gehrke, and Lars Vilhuber. Privacy: Theory meets practice on the map. In *2008 IEEE 24th International Conference on Data Engineering*, pages 277–286. IEEE, 2008.

[32] Charles F Manski. The lure of incredible certitude. *Economics & Philosophy*, 36(2):216–245, 2020.

[33] David McClure and Jerome P Reiter. Differential privacy and statistical disclosure risk measures: An investigation with binary synthetic data. *Transactions on Data Privacy*, 5(3):535–552, 2012.

[34] Luise Mehner, Saskia Nuñez von Voigt, and Florian Tschorsch. Towards explaining epsilon: A worst-case study of differential privacy risks. In *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 328–331. IEEE, 2021.

[35] Priyanka Nanayakkara, Johes Bater, Xi He, Jessica Hullman, and Jennie Rogers. Visualizing privacy-utility trade-offs in differentially private data releases. *Proceedings on Privacy Enhancing Technologies*, 2022.

[36] Priyanka Nanayakkara, Mary Anne Smart, Rachel Cummings, Gabriel Kaptchuk, and Elissa M Redmiles. What are the chances? explaining the epsilon parameter in differential privacy. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 1613–1630, 2023.

[37] NCDHHS. 2020 north carolina infant mortality report, table 1. NC State Center for Health Statistics, 2020. Accessed on 22 April 2024.

[38] NCDHHS. Cy2020 north carolina resident births by county of residence and month. NC State Center for Health Statistics, 2020. Accessed on 22 April 2024.

[39] Alisa Pankova and Peeter Laud. Interpreting epsilon of differential privacy in terms of advantage in guessing or approximating sensitive attributes. In *2022 IEEE 35th Computer Security Foundations Symposium (CSF)*, pages 96–111. IEEE, 2022.

[40] Giovanni Parmigiani and Lurdes Inoue. *Decision theory: Principles and approaches*. John Wiley & Sons, 2009.

[41] Jerome P Reiter. Estimating risks of identification disclosure in microdata. *Journal of the American Statistical Association*, 100(472):1103–1112, 2005.

[42] Jerome P Reiter. Differential privacy and federal data releases. *Annual Review of Statistics and Its Application*, 6:85–101, 2019.

[43] Jayshree Sarathy, Sophia Song, Audrey Haque, Tania Schlatter, and Salil Vadhan. Don't look at the data! how differential privacy reconfigures the practices of data science. *arXiv preprint arXiv:2302.11775*, 2023.

[44] Jun Tang, Aleksandra Korolova, Xiaolong Bai, Xueqiang Wang, and Xiaofeng Wang. Privacy loss in Apple's implementation of differential privacy on macos 10.12. *arXiv preprint arXiv:1709.02753*, 2017.

[45] Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R O'Brien, Thomas Steinke, and Salil Vadhan. Differential privacy: A primer for a non-technical audience. *Vanderbilt Journal of Entertainment & Technology Law*, 21:209, 2018.

[46] Alexandra Wood, Micah Altman, Kobbi Nissim, and Salil Vadhan. Designing access with differential privacy. In *Handbook on Using Administrative Data for Research and Evidence-based Policy*. Abdul Latif Jameel Poverty Action Lab, Cambridge, MA, 2020.

A Supplemental Tables

A.1 Notation Summary

Table 2 summarizes the notation of Section 2. Table 3 summarizes the definitions of key probabilistic quantities.

Symbol	Description
$\mathbf{P}$	Population of individuals the data is drawn from
$\mathbf{Y}$	$n \times d$ confidential data set
Y_i	Length-d vector of values for individual i
I_i	Indicator for whether individual i is included in $\mathbf{Y}$
$\mathbf{Y}_{-i}$	$(n-1) \times d$ matrix of the values in $\mathbf{Y}$ excluding those for individual i
$\mathcal{S}$	Subset of the support of Y_i constituting a privacy violation
$r^*(p_i, q_i)$	Function describing the agency's desired relative risk bound
$\mathcal{M}$	Adversary's model for predicting Y_i
T^*	Noisy estimate of $T(\mathbf{Y})$, the function being released

Table 2: Summary of notation.

Symbol	Definition	Description
p_i	$P_{\mathcal{M}}[I_i = 1]$	Prior probability of inclusion
q_i	$P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid I_i = 1]$	Prior probability values disclosed
$r_i(p_i, q_i, t^*)$	$\frac{P_{\mathcal{M}}[Y_i \in \mathcal{S}, I_i = 1 \mid T^* = t^*]}{P_{\mathcal{M}}[Y_i \in \mathcal{S}, I_i = 1]}$	Relative disclosure risk
$a_i(p_i, q_i, t^*)$	$P_{\mathcal{M}}[Y_i \in \mathcal{S}, I_i = 1 \mid T^* = t^*]$	Absolute disclosure risk

Table 3: Summary of definitions.

A.2 Closed Forms for ε

In this section, we provide a reference for all derived closed forms for ε. Corresponding formal theorem statements are provided in Appendix B.1 and discussion of closed forms not referenced in the main text is provided in Appendix B.2. Table 4 provides the closed forms.

Risk Profile	Condition	Recommended ε
(14); $\tilde{r}$		$\frac{1}{2}\log(\tilde{r})$
	$0 < \tilde{q} \leq \frac{\tilde{a}}{\tilde{r}}$	$\frac{1}{2}\log\left(\frac{\tilde{a}(1-\tilde{q})}{\tilde{q}(1-\tilde{a})}\right)$
(15)	$0 < \tilde{q} \leq \frac{1}{\tilde{r}+1}$ and $\frac{\tilde{a}}{\tilde{r}} < \tilde{q} < 1$	$\frac{1}{2}\log\left(\frac{1-\tilde{q}}{\frac{1}{\tilde{r}}-\tilde{q}}\right)$
$\max\left\{\frac{\tilde{a}}{p_i\tilde{q}}, \tilde{r}\right\}$	$\frac{1}{\tilde{r}+1} < \tilde{q} < 1$ and $\frac{\tilde{a}}{\tilde{r}} < \tilde{q} < 1$	$\log\left(\frac{2\tilde{a}(1-\tilde{q})}{\sqrt{(\tilde{r}\tilde{q}-\tilde{a})^2+4\tilde{a}\tilde{q}(1-\tilde{q})(1-\tilde{a})}-(\tilde{r}\tilde{q}-\tilde{a})}\right)$
	$\tilde{q} = 1$	$\log\left(\frac{\tilde{r}-\tilde{a}}{1-\tilde{a}}\right)$
(16)	$0 < \tilde{p} \leq \frac{\tilde{a}}{\tilde{r}}$	$\log\left(\frac{\tilde{a}(1-\tilde{p})}{\tilde{p}(1-\tilde{a})}\right)$
$\max\left\{\frac{\tilde{a}}{\tilde{p}q_i}, \tilde{r}\right\}$	$\frac{\tilde{a}}{\tilde{r}} < \tilde{p} \leq 1$	$\log\left(\frac{2(\tilde{p}\tilde{r}-\tilde{a})}{\sqrt{\tilde{r}^2(1-\tilde{p})^2+4(\tilde{p}\tilde{r}-\tilde{a})(1-\tilde{a})}-\tilde{r}(1-\tilde{p})}\right)$
(25)	$0 \leq \tilde{q}_0 \leq \frac{1}{\tilde{r}+1}$	$\log\left(\frac{2\tilde{p}_1(1-\tilde{q}_0)}{\sqrt{(1-\tilde{p}_1)^2+4\tilde{p}_1(1-\tilde{q}_0)\left(\frac{1}{\tilde{r}}-\tilde{p}_1\tilde{q}_0\right)}-(1-\tilde{p}_1)}\right)$
$\tilde{r}$	$\frac{1}{\tilde{r}+1} < \tilde{q}_0 < 1$ and $\tilde{p}_0 > 0$	$\log\left(\frac{2\tilde{p}_0(1-\tilde{q}_0)}{\sqrt{(1-\tilde{p}_0)^2+4\tilde{p}_0(1-\tilde{q}_0)\left(\frac{1}{\tilde{r}}-\tilde{p}_0\tilde{q}_0\right)}-(1-\tilde{p}_0)}\right)$
if $\tilde{p}_0 \leq p_i \leq \tilde{p}_1$	$\frac{1}{\tilde{r}+1} < \tilde{q}_0 < 1$ and $\tilde{p}_0 = 0$	$\log(\tilde{r})$
and $\tilde{q}_0 \leq q_i \leq \tilde{q}_1$	$\tilde{q}_0 = 1$	$\log\left(\frac{1-\tilde{p}_0}{\frac{1}{\tilde{r}}-\tilde{p}_0}\right)$
(28); $\frac{p_iq_i+\tilde{b}}{p_iq_i}$		$\log\left(\frac{1+\tilde{b}}{1-\tilde{b}}\right)$

Table 4: Closed forms for ε for various $r^*(p_i, q_i)$.

B Additional Results

B.1 Omitted Results

In this section, we present all results omitted from the paper. Proofs of these results are included in Appendix E.

First, we state a corollary to Theorem 1, which generalizes Theorem 1.3 in [19] to sets where $|\mathcal{S}| > 1$ and to unbounded DP.

Corollary 1. *Under the conditions of Theorem 1, for all $p_i, q_i \in (0, 1]$ and all t^*,*

$$r_i(p_i, q_i, t^*) \leq e^{2\varepsilon}. \tag{18}$$

Next, we state a corollary to Theorem 2 which considers the special case where $p_i = 1$. This corresponds to a setting where individual i's inclusion in the data is known a priori by the adversary, for example data from a census or public social media platform.

Corollary 2. *Under the conditions of Theorem 2, if $p_i = 1$ and $0 < q_i < 1$, any statistic $T^* = t^*$ released under ε-DP with*

$$\varepsilon \leq \frac{1}{2}\log\left(\frac{1 - q_i}{\frac{1}{r^*(1,q_i)} - q_i}\right), \tag{19}$$

will satisfy $r_i(1, q_i, t^*) \leq r^*(1, q_i)$.

For particular forms of r^*, the optimization in (13) has a closed form solution. This is detailed by the following theorems, which are preceded by two lemmas used in the proofs.

Lemma 2. *Fix $p_i, q_i \in (0, 1)$ and let $r^*(p_i, q_i) = \frac{\tilde{a}}{p_i q_i}$. Then the function*

$$\varepsilon(p_i, q_i) = \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_i q_i\right)} - (1-p_i)}\right) \tag{20}$$

has partial derivatives such that

1. $\frac{\partial\varepsilon(p_i,q_i)}{\partial p_i} < 0$ *for all* $0 < p_i < 1$ *and* $0 < q_i < 1$
2. $\frac{\partial\varepsilon(p_i,q_i)}{\partial q_i} < 0$ *for all* $0 < p_i < 1$ *and* $0 < q_i < 1$.

Lemma 3. *Fix $p_i, q_i \in (0, 1)$ and let $r^*(p_i, q_i) = \tilde{r}$. Then the function*

$$\varepsilon(p_i, q_i) = \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_i q_i\right)} - (1-p_i)}\right) \tag{21}$$

has partial derivatives such that

1. $\frac{\partial\varepsilon(p_i,q_i)}{\partial p_i} < 0$ *if* $0 < q_i < \frac{1}{\tilde{r}+1}$ *and* $0 < p_i < 1$
2. $\frac{\partial\varepsilon(p_i,q_i)}{\partial p_i} = 0$ *if* $q_i = \frac{1}{\tilde{r}+1}$ *and* $0 < p_i < 1$
3. $\frac{\partial\varepsilon(p_i,q_i)}{\partial p_i} > 0$ *if* $\frac{1}{\tilde{r}+1} < q_i < 1$ *and* $0 < p_i < 1$
4. $\frac{\partial\varepsilon(p_i,q_i)}{\partial q_i} > 0$ *for all* $0 < p_i < 1$ *and* $0 < q_i < 1$.

Theorem 3. *Under the conditions of Theorem 2, if $r^*(p_i, q_i) = \tilde{r} > 1$, the solution to the minimization problem in (13) is*

$$\varepsilon = \frac{1}{2}\log\left(\tilde{r}\right). \tag{22}$$

Theorem 4. *Under the conditions of Theorem 2, let $\tilde{a} < 1$, $\tilde{p} \leq 1$, and $\tilde{r} > 1$, and $0 < q_i < 1$. If the function r^* is such that $r^*(\tilde{p}, q_i) = \max\{\tilde{a}/(\tilde{p}q_i), \tilde{r}\}$ and $r^*(p_i, q_i) = \infty$ if $p_i \neq \tilde{p}$, then the solution to the minimization problem in (13) is*

$$\varepsilon = \begin{cases} \log\left(\frac{\tilde{a}(1-\tilde{p})}{\tilde{p}(1-\tilde{a})}\right), & \textit{if } 0 < \tilde{p} \leq \frac{\tilde{a}}{\tilde{r}}; \\ \log\left(\frac{2(\tilde{p}\tilde{r}-\tilde{a})}{\sqrt{\tilde{r}^2(1-\tilde{p})^2+4(\tilde{p}\tilde{r}-\tilde{a})(1-\tilde{a})}-\tilde{r}(1-\tilde{p})}\right), & \textit{if } \frac{\tilde{a}}{\tilde{r}} < \tilde{p} \leq 1. \end{cases} \tag{23}$$

Theorem 5. *Under the conditions of Theorem 2, let $\tilde{a} < 1$, $\tilde{q} \leq 1$, and $\tilde{r} > 1$, and $0 < p_i < 1$. If the function r^* is such that $r^*(p_i, \tilde{q}) = \max\{\tilde{a}/(p_i\tilde{q}), \tilde{r}\}$ and $r^*(p_i, q_i) = \infty$ for $q_i \neq \tilde{q}$, then the solution to the minimization problem in (13) is*

$$\varepsilon = \begin{cases} \frac{1}{2}\log\left(\frac{\tilde{a}(1-\tilde{q})}{\tilde{q}(1-\tilde{a})}\right), & \textit{if } 0 < \tilde{q} \leq \frac{\tilde{a}}{\tilde{r}}; \\ \frac{1}{2}\log\left(\frac{1-\tilde{q}}{\frac{1}{\tilde{r}}-\tilde{q}}\right), & \textit{if } 0 < \tilde{q} \leq \frac{1}{\tilde{r}+1} \textit{ and } \frac{\tilde{a}}{\tilde{r}} < \tilde{q} < 1; \\ \log\left(\frac{2\tilde{a}(1-\tilde{q})}{\sqrt{(\tilde{r}\tilde{q}-\tilde{a})^2+4\tilde{a}\tilde{q}(1-\tilde{q})(1-\tilde{a})}-(\tilde{r}\tilde{q}-\tilde{a})}\right), & \textit{if } \frac{1}{\tilde{r}+1} < \tilde{q} < 1 \textit{ and } \frac{\tilde{a}}{\tilde{r}} < \tilde{q} < 1; \\ \log\left(\frac{\tilde{r}-\tilde{a}}{1-\tilde{a}}\right), & \textit{if } \tilde{q} = 1. \end{cases} \tag{24}$$

B.2 Additional Closed Forms

In this section, we present closed form results for the ε implied by additional disclosure risk profiles not discussed in the main text. Proofs of these results are not provided, although we note that proofs would have a structure similar to the proofs of Theorems 3–5. It is straightforward to verify these results empirically.

First, consider an agency that requires a constant relative risk bound to hold on a subset of the (p_i, q_i) space and does not have any bounds outside that space. That is, rather than enforcing $r^*(p_i, q_i) = \tilde{r}$ for all $0 \leq p_i, q_i \leq 1$, the agency only enforces this condition for $\tilde{p}_0 \leq p_i \leq \tilde{p}_1$ and $\tilde{q}_0 \leq q_i \leq \tilde{q}_1$, where $0 \leq \tilde{p}_0 \leq \tilde{p}_1 \leq 1$ and $0 \leq \tilde{q}_0 \leq \tilde{q}_1 \leq 1$ (for $\tilde{p}_1, \tilde{q}_1 > 0$). Formally,

$$r^*(p_i, q_i) = \begin{cases} \tilde{r}, & \text{if } \tilde{p}_0 \leq p_i \leq \tilde{p}_1 \text{ and } \tilde{q}_0 \leq q_i \leq \tilde{q}_1; \\ \infty, & \text{otherwise.} \end{cases} \tag{25}$$

The recommended ε for an agency with the risk profile in (25) is

$$\varepsilon = \begin{cases} \log\left(\frac{2\tilde{p}_1(1-\tilde{q}_0)}{\sqrt{(1-\tilde{p}_1)^2+4\tilde{p}_1(1-\tilde{q}_0)\left(\frac{1}{\tilde{r}}-\tilde{p}_1\tilde{q}_0\right)}-(1-\tilde{p}_1)}\right), & \text{if } 0 \leq \tilde{q}_0 \leq \frac{1}{\tilde{r}+1}; \\ \log\left(\frac{2\tilde{p}_0(1-\tilde{q}_0)}{\sqrt{(1-\tilde{p}_0)^2+4\tilde{p}_0(1-\tilde{q}_0)\left(\frac{1}{\tilde{r}}-\tilde{p}_0\tilde{q}_0\right)}-(1-\tilde{p}_0)}\right), & \text{if } \frac{1}{\tilde{r}+1} < \tilde{q}_0 < 1 \text{ and } \tilde{p}_0 > 0; \\ \log(\tilde{r}), & \text{if } \frac{1}{\tilde{r}+1} < \tilde{q}_0 < 1 \text{ and } \tilde{p}_0 = 0; \\ \log\left(\frac{1-\tilde{p}_0}{\frac{1}{\tilde{r}}-\tilde{p}_0}\right), & \text{if } \tilde{q}_0 = 1. \end{cases} \tag{26}$$

Next, consider an agency that desires a constant bound on the difference between the posterior and the prior. That is, for some $\tilde{b} \in (0, 1)$, they would like to enforce

$$P_{\mathcal{M}}[Y_i \in \mathcal{S}, I_i = 1 \mid T^* = t^*] - P_{\mathcal{M}}[Y_i \in \mathcal{S}, I_i = 1] \leq \tilde{b}. \tag{27}$$

This constraint corresponds to the following risk profile.

$$r^*(p_i, q_i) = \frac{p_i q_i + \tilde{b}}{p_i q_i} = 1 + \frac{\tilde{b}}{p_i q_i}. \tag{28}$$

The recommended ε for an agency with this risk profile is

$$\varepsilon = \log\left(\frac{1+\tilde{b}}{1-\tilde{b}}\right). \tag{29}$$

Notably, by a Taylor series approximation, $-\log(1-\tilde{b}) \approx \tilde{b} + \tilde{b}^2/2$. Thus, for small $\tilde{b}$, it follows that the recommended ε can be approximated by

$$\varepsilon = \log(1+\tilde{b}) - \log(1-\tilde{b}) \approx \left(\tilde{b} - \frac{\tilde{b}^2}{2}\right) + \left(\tilde{b} + \frac{\tilde{b}^2}{2}\right) = 2\tilde{b}. \tag{30}$$

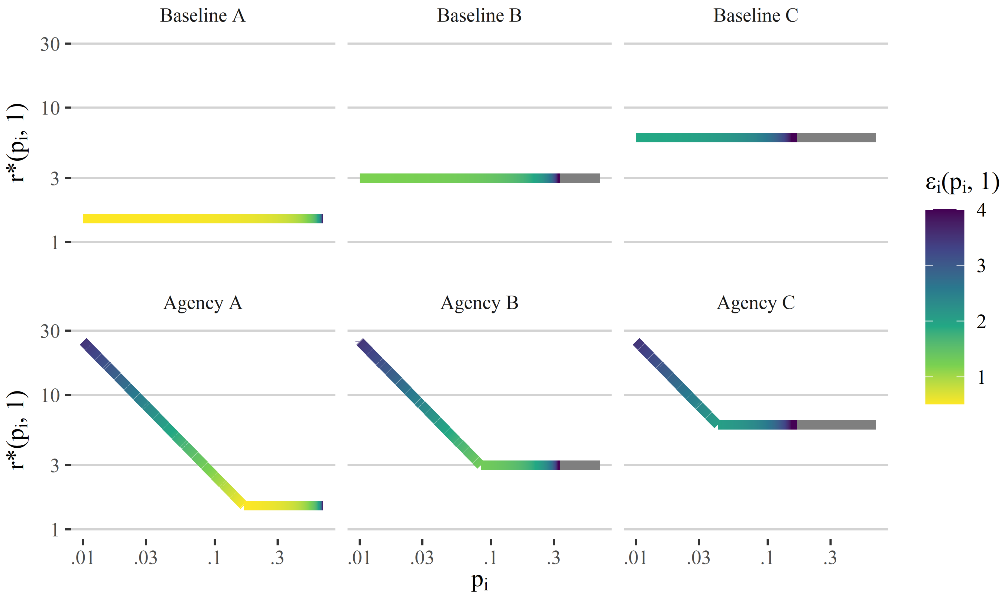

Figure 3: The risk profiles for three agencies with risk profile given by (31). The lines in the lower panels represent the risk profiles for $q_i = 1$ as a function of p_i, and the colors represent the implied ε at each point on the curve. The lines in the upper panels represent the corresponding baseline $r^*(p_i, 1) = \tilde{r}$. The left plots set $\tilde{r} = 1.5$, the center plots set $\tilde{r} = 3$, and the right plots set $\tilde{r} = 6$.

C Additional Examples

This section provides additional examples of risk profiles.

C.1 Additional Figure for Example 1

Here we provide an additional figure for Example 1. The setting is restated below.

Example 1. *A healthcare provider possesses a data set comprising demographic information about individuals diagnosed with a particular form of cancer in a region of interest. They plan to release the count of individuals diagnosed with this form of cancer in various demographic groups via the geometric mechanism, but are concerned this release, if insufficient noise is added, could reveal which individuals in the community have cancer. They wish to choose ε appropriately.*

As discussed in Section 4, a reasonable risk profile for this setting might focus on p_i and set $q_i = 1$. If we additionally suppose the agency is generally willing to accept a maximum absolute disclosure risk of 0.25 for adversaries with small prior probabilities. A risk profile for this agency might be of the form, for some $\tilde{r} > 1$,

$$r^*(p_i, q_i) = \begin{cases} \max\left\{\frac{0.25}{p_i}, \tilde{r}\right\}, & \text{if } q_i = 1; \\ \infty, & \text{if } q_i \neq 1. \end{cases} \tag{31}$$

The three example risk functions considered in Section 4 are presented in Figure 3. Agency A is risk averse, agency C is utility seeking, and agency B sits in between in terms of risk and utility. For adversaries with high prior probabilities, agencies A, B, and C bound the relative disclosure risk at $\tilde{r} = 1.5$, $\tilde{r} = 3$, and $\tilde{r} = 6$, respectively.

C.2 Extended Example 2

Here we provide an extended analysis of Example 2. The setting is restated below.

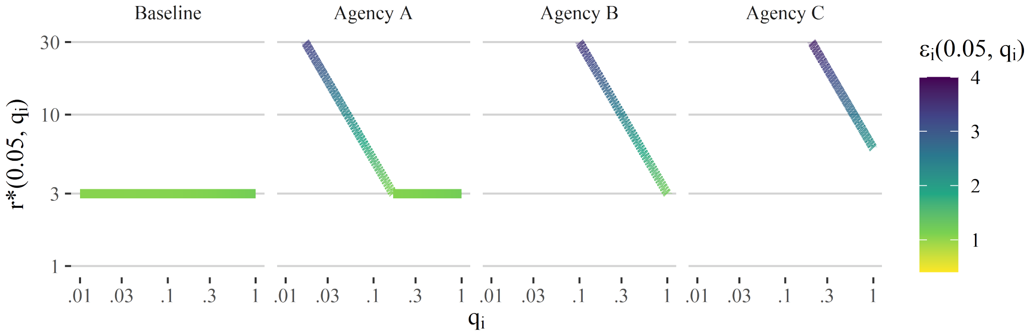

Figure 4: The risk profiles for three agencies with risk profile given by (32). The lines represent the risk profiles for $p_i = 0.05$ as a function of q_i, and the colors represent the implied ε at each point on the curve. The left plot sets $\tilde{a} = 0.025$, the center sets $\tilde{a} = 0.15$, and the right sets $\tilde{a} = 0.3$.

Agency	$\tilde{a}$	ε Recommendation	Noise Std. Dev.	Prob. Exact
A	0.025	1.09	1.24	50%
B	0.15	1.21	1.10	54%
C	0.3	2.10	0.56	78%

Table 5: For each of the three risk profiles in Figure 4, we present the ε recommended by our framework. For a release satisfying ε-DP using the geometric mechanism, we present the corresponding standard deviation of the noise distribution and the probability that the exact value is released (i.e., the noise distribution's probability mass at zero).

Example 2. *A survey is performed on a sample of individuals in the region of interest. 5% of the region is surveyed, and respondents are asked whether they have this particular form of cancer along with a series of demographic questions. The agency plans to release the counts of surveyed individuals who have and do not have the cancer in various demographic groups via the geometric mechanism, but is concerned this release, if insufficient noise is added, could reveal which individuals in the community have cancer. The agency wishes to choose ε appropriately.*

As discussed in Section 4, a reasonable risk profile for this setting might set $\mathcal{S} = \{y \in \{0,1\}^d : y_1 = 1\}$ and focus on q_i, while fixing $p_i = 0.05$. Additionally, suppose the agency is generally willing to accept a maximum relative disclosure risk of 3 for adversaries with large prior probabilities. A reasonable risk profile for this agency might be of the form, for some $\tilde{a} < 1$,

$$r^*(p_i, q_i) = \begin{cases} \max\left\{\frac{\tilde{a}}{0.05 q_i}, 3\right\}, & \text{if } p_i = 0.05; \\ \infty, & \text{if } p_i \neq 0.05. \end{cases} \tag{32}$$

Three example risk functions of this form are presented in Figure 4. Once again, agency A is risk averse, agency C is utility seeking, and agency B sits in between on risk and utility. Agencies A, B, and C are willing to allow adversaries to achieve an absolute disclosure risk of $\tilde{a} = 0.025$, $\tilde{a} = 0.15$, and $\tilde{a} = 0.3$, respectively. Table 5 presents the ε recommendations for each agency along with the standard deviation of the noise and probability of releasing the exact value of each statistic under the geometric mechanism.

As in Table 1, the ε recommendations reflect trade offs between privacy and accuracy. They also are much higher than the recommendations from a corresponding simple risk profile of $r^*(p_i, q_i) = 3$ for all prior probabilities, which implies $\varepsilon \approx 0.55$. Even the most risk averse agency is recommended an ε that is much larger than this baseline risk profile. This gain is primarily due to the assumptions that the survey is a simple random sample from the population, and the adversary has no prior knowledge about which individuals are surveyed. Essentially, the additional uncertainty from the sampling mechanism allows for an ε recommendation with less noise injected. This is consistent with prior work showing that DP mechanisms applied to random subsamples provide better privacy guarantees [3]. The recommended ε will increase as the proportion of individuals from the population surveyed

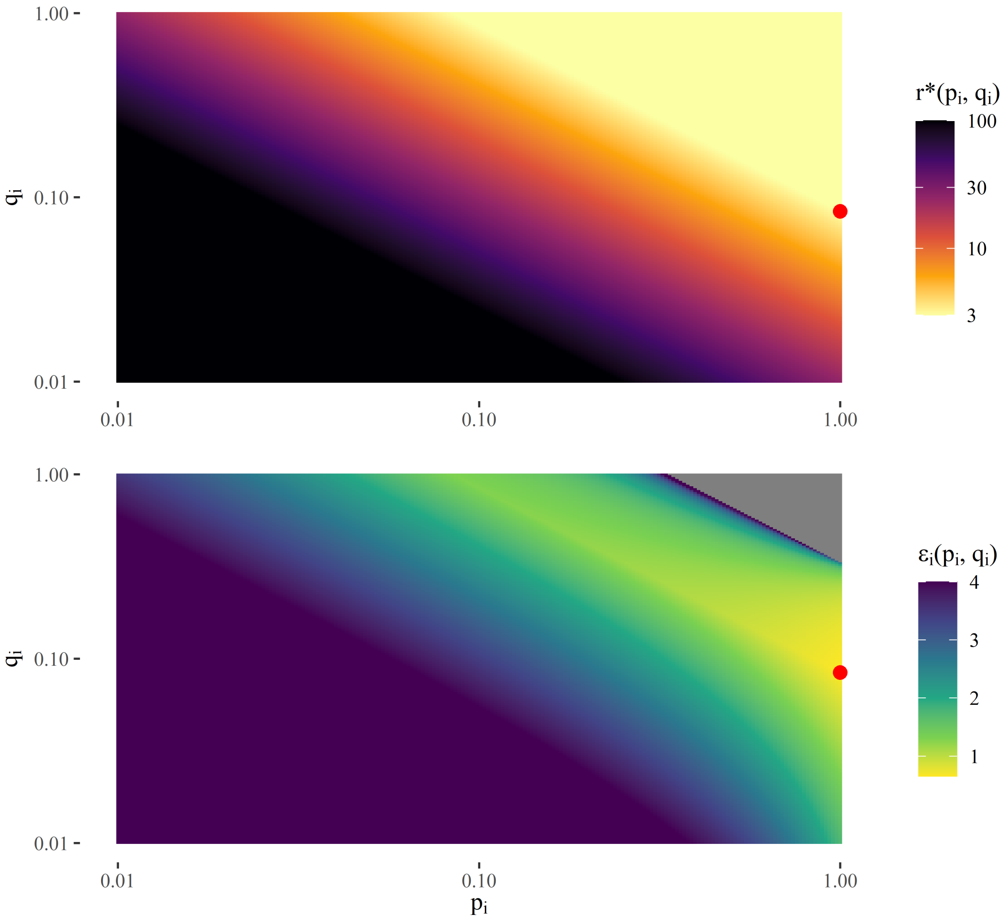

Figure 5: The top panel presents the $r^*(p_i, q_i)$ from (33) as a function of p_i and q_i. The bottom panel presents the implied $\varepsilon_i(p_i, q_i)$ as a function of p_i and q_i. The red point represents $\text{argmin}_{(p_i,q_i)}\, \varepsilon_i(p_i, q_i)$. For clarity of presentation, all $r^*(p_i, q_i) > 100$ are truncated to 100 and all $\varepsilon_i(p_i, q_i) > 4$ are truncated to 4.

($\tilde{p}$) decreases. For example, for the risk function in (32) with $\tilde{a} = 0.025$ and $\tilde{r} = 3$, if $\tilde{p} = 0.005$, we recommend $\varepsilon \approx 1.63$; if $\tilde{p} = 0.0005$, we recommend $\varepsilon \approx 3.94$.

C.3 Two-dimensional Example

The examples in Appendices C.1 and C.2 focus on settings where only one of p_i and q_i is important and the other can be reasonably fixed to a constant. Now we consider the setting where both may be simultaneously important. For example, in Example 2, the agency may wish to consider adversaries with various p_i, rather than only $\tilde{p} = 0.05$. Perhaps the agency wishes to limit the absolute disclosure to be below $\tilde{a} = 0.25$ when total prior probability of disclosure, $p_i q_i$, is low and limit relative disclosure to be below $\tilde{r} = 3$ when $p_i q_i$ is high. This corresponds to the following risk profile.

$$r^*(p_i, q_i) = \max\left\{\frac{\tilde{a}}{p_i q_i}, \tilde{r}\right\} = \max\left\{\frac{0.25}{p_i q_i}, 3\right\}. \tag{33}$$

A plot of this desired bound on pairs (p_i, q_i) on the space $(0, 1] \times (0, 1]$ is presented in the top panel of Figure 5. Notably, if p_i is fixed at $\tilde{p} \in (0, 1]$, a plot of $r^*(\tilde{p}, q_i)$ as a function of q_i has a form similar to the plots of Figure 4. Similarly, if q_i is fixed at $\tilde{q} \in (0, 1]$, a plot of $r^*(p_i, \tilde{q})$ as a function of p_i has a form similar to the plots of Figure 3.

To determine the ε recommended by this profile, we numerically solve the minimization problem in (13). A plot of $\varepsilon_i(p_i, q_i)$ as a function of pairs (p_i, q_i) on the space $(0, 1] \times (0, 1]$ is presented in the bottom panel of Figure 5. Our framework recommends $\varepsilon \approx 0.65$, which corresponds to $p_i = 1$ and $q_i = 0.083 = \tilde{a}/\tilde{r}$.

D Additional Discussion of Prior Work

In this section, we describe two previous works with similar aims to our framework using the notation in Section 2.2, expanding on the discussion in Section 6.

D.1 “How Much is Enough? Choosing ε for Differential Privacy"

[29] focus on settings where the population, $\mathbf{P}$, of size n is public information and the adversary's goal is to determine which subset of individuals in $\mathbf{P}$ was used for a differentially private release of a statistic. We can characterize their setting with the notation of Section 2.2 as follows. We define $\mathbf{Y}$ to be the subset of individuals' values in $\mathbf{P}$ used to compute the statistic of interest, $T(\mathbf{Y})$, and its released DP counterpart, $T^*(\mathbf{Y})$. In their examples, the authors focus on the setting where only one individual is removed from $\mathbf{P}$ to create $\mathbf{Y}$, and the adversary's goal is to determine which i was removed.

We can apply our framework to this setting with a minor modification. For this comparison, we assume the adversary's $q_i = P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid I_i = 0] = 1$ for any set $\mathcal{S}$ (although we note that this is a weaker assumption than that of Lee and Clifton, since we do not assume $\mathbf{P}$ is public). We redefine p_i and the risk measures to be in terms of $I_i = 0$, rather than $I_i = 1$.

$$p_i = P_{\mathcal{M}}[I_i = 0] \tag{34}$$

$$r_i(p_i, 1, t^*) = \frac{P_{\mathcal{M}}[I_i = 0 \mid T^* = t^*]}{P_{\mathcal{M}}[I_i = 0]} \tag{35}$$

$$a_i(p_i, 1, t^*) = P_{\mathcal{M}}[I_i = 0 \mid T^* = t^*]. \tag{36}$$

[29] focus on the case of $p_i = 1/n$, and seek to enforce the bound $a_i(1/n, 1, t^*) \leq \tilde{a}$ for some constant $\tilde{a}$ and all t^*, which implies the relative risk bound

$$r^*(p_i, q_i) = \begin{cases} n\tilde{a}, & \text{if } p_i = \frac{1}{n}, q_i = 1; \\ \infty, & \text{otherwise.} \end{cases} \tag{37}$$

From an analogy to Theorem 2 with the redefined p_i, it follows that under these conditions, our method sets

$$\varepsilon = \log\left(\frac{1 - \frac{1}{n}}{\frac{1}{n\tilde{a}} - \frac{1}{n}}\right) = \log\left(\frac{(n-1)\tilde{a}}{1 - \tilde{a}}\right). \tag{38}$$

In the motivating example from their paper, the authors set $n = 4$ and $\tilde{a} = 1/3$, giving $r^*(1/n, 1) = 4/3$. This results in $\varepsilon = \log(3/2) \approx 0.41$ from our method. When the release mechanism is the addition of Laplace noise, [29]'s method arrives at a similar form, but with the recommendation scaled by a factor of $\Delta T/\Delta v$.

$$\varepsilon = \frac{\Delta T}{\Delta v} \log\left(\frac{(n-1)\tilde{a}}{1 - \tilde{a}}\right), \tag{39}$$

$$\Delta T = \max\left\{|T(\mathbf{Y}_1) - T(\mathbf{Y}_2)| : \mathbf{Y}_2 \subset \mathbf{Y}_1 \subset \mathbf{P}, |\mathbf{Y}_1| = n-1, |\mathbf{Y}_2| = n-2\right\} \tag{40}$$

$$\Delta v = \max\left\{|T(\mathbf{Y}_1) - T(\mathbf{Y}_2)| : \mathbf{Y}_1 \subset \mathbf{P}, \mathbf{Y}_2 \subset \mathbf{P}, |\mathbf{Y}_1| = |\mathbf{Y}_2| = n-1\right\}. \tag{41}$$

The recommendation of [29]'s method thus depends on both the population and the particular release function. This results in different ε values for the same n and $\tilde{a}$, as low as 0.34 and as high as 1.62 in the authors' examples, depending on the statistic of interest and the values in the data.

D.2 “Differential Privacy: A Primer for a Non-Technical Audience"

The example in [45] considers an individual deciding whether or not to participate in a survey for which results will be released via DP with a particular ε. Using our notation, let $Z_i = f(Y_i, I_i) \in$

$\{0, 1\}$ be a quantity of interest to the adversary, who wishes to learn whether $Z_i = 1$. They have some prior $q_i = P_{\mathcal{M}}[Z_i = 1]$. Rather than considering the relative or absolute disclosure risk, the individual is interested in comparing the adversary's posterior probability if they participate in the survey, $a_{1i}(q_i, t^*) = P_{\mathcal{M}}[Z_i = 1 \mid I_i = 1, T^* = t^*]$, to the adversary's posterior probability if they do not participate, $a_{0i}(q_i, t^*) = P_{\mathcal{M}}[Z_i = 1 \mid I_i = 0, T^* = t^*]$. The authors of [45] state that for all q_i and all t^*,

$$a_{1i}(q_i, t^*) \leq \frac{a_{0i}(q_i, t^*)}{a_{0i}(q_i, t^*) + e^{-2\varepsilon}(1 - a_{0i}(q_i, t^*))}. \tag{42}$$

This expression is in the same spirit as the results from our framework with $p_i = 1$. By Theorem 1, we have

$$r_i(1, q_i, t^*) \leq \frac{1}{q_i + e^{-2\varepsilon}(1 - q_i)} \quad \Longrightarrow \quad a_i(1, q_i, t^*) \leq \frac{q_i}{q_i + e^{-2\varepsilon}(1 - q_i)}. \tag{43}$$

The authors of [45] suggest that the individual considering survey participation use (42) to bound a_{1i} for various values of a_{0i}. The individual can examine these bounds to make an informed decision about whether to participate in the survey.

E Proofs

This section provides proofs of results from Section 3 and Appendix B.1.

Lemma 1. *Under Assumption 1 and Assumption 2, if the release of $T^* = t^*$ satisfies DP, then for any subset $\mathcal{S}$ of the domain of Y_i, we have*

$$e^{-\varepsilon} \leq \frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} \leq e^{\varepsilon}, \qquad e^{-2\varepsilon} \leq \frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1]} \leq e^{2\varepsilon}. \tag{10}$$

Proof. To begin, we demonstrate that for all y in the support of Y_i and t^* in the support of T^*,

$$e^{-\varepsilon} \leq \frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i = y, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} \leq e^{\varepsilon}. \tag{44}$$

Let $\mathcal{Y}_{-i}$ be the support of $\mathbf{Y}_{-i}$ under $\mathcal{M}$. Then,

$$\begin{aligned}
&P_{\mathcal{M}}[T^* = t^* \mid Y_i = y, I_i = 1] \\
&\quad = \sum_{\mathbf{y}_{-i} \in \mathcal{Y}_{-i}} P[T^* = t^* \mid Y_i = y, \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 1]\, P_{\mathcal{M}}[\mathbf{Y}_{-i} = \mathbf{y}_{-i} \mid Y_i = y, I_i = 1] && (45)\\
&\quad \leq \sum_{\mathbf{y}_{-i} \in \mathcal{Y}_{-i}} e^{\varepsilon} P[T^* = t^* \mid \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 0]\, P_{\mathcal{M}}[\mathbf{Y}_{-i} = \mathbf{y}_{-i} \mid Y_i = y, I_i = 1] && (46)\\
&\quad = e^{\varepsilon} \sum_{\mathbf{y}_{-i} \in \mathcal{Y}_{-i}} P_{\mathcal{M}}[T^* = t^* \mid \mathbf{Y}_{-i} = \mathbf{y}_{-i}, I_i = 0]\, P[\mathbf{Y}_{-i} = \mathbf{y}_{-i} \mid I_i = 0] && (47)\\
&\quad = e^{\varepsilon} P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]. && (48)
\end{aligned}$$

The equality in (45) follows from the law of total probability and Assumption 1. The inequality in (46) follows from DP via (1). The equality in (47) follows from Assumption 2. The equality in (48) follows from the law of total probability and Assumption 1. This completes the proof of the right inequality. The proof of the left inequality is identical with the other inequality in (1) applied in (46).

Now, we apply Bayes' Theorem to $P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]$.

$$\begin{aligned}
\frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} &= \frac{\frac{P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid T^* = t^*, I_i = 1]\, P_{\mathcal{M}}[T^* = t^* \mid I_i = 1]}{P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid I_i = 1]}}{P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} && (49)\\
&= \frac{P_{\mathcal{M}}[T^* = t^* \mid I_i = 1]\, P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid T^* = t^*, I_i = 1]}{P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid I_i = 1]\, P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} && (50)
\end{aligned}$$

We can then break the second term in the numerator into a summation and apply Bayes' Theorem to each term in the sum.

$$\begin{aligned}
&\frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} \\
&\qquad = \frac{P_{\mathcal{M}}[T^* = t^* \mid I_i = 1] \sum_{y \in \mathcal{S}} P_{\mathcal{M}}[Y_i = y \mid T^* = t^*, I_i = 1]}{P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid I_i = 1]\, P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} && (51)\\
&\qquad = \frac{P_{\mathcal{M}}[T^* = t^* \mid I_i = 1] \sum_{y \in \mathcal{S}} \frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i = y, I_i = 1]\, P_{\mathcal{M}}[Y_i = y \mid I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid I_i = 1]}}{P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid I_i = 1]\, P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} && (52)\\
&\qquad = \frac{\sum_{y \in \mathcal{S}} \frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i = y, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} P_{\mathcal{M}}[Y_i = y \mid I_i = 1]}{P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid I_i = 1]} && (53)
\end{aligned}$$

We apply (44) to achieve the left bound in the first expression of (10).

$$\frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} \geq \frac{\sum_{y \in \mathcal{S}} e^{-\varepsilon} P_{\mathcal{M}}[Y_i = y \mid I_i = 1]}{P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid I_i = 1]} = e^{-\varepsilon}. \tag{54}$$

We achieve the right bound in the first expression of (10) in a similar fashion.

$$\frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} \leq \frac{\sum_{y \in \mathcal{S}} e^{\varepsilon} P_{\mathcal{M}}[Y_i = y \mid I_i = 1]}{P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid I_i = 1]} = e^{\varepsilon}. \tag{55}$$

We now turn to the second expression in (10). First note that

$$\begin{aligned}&\frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1]}\\ &\qquad = \frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} \cdot \frac{P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]}{P_{\mathcal{M}}[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1]}. \qquad (56)\end{aligned}$$

By applying the bounds in the first expression in (10) to $\mathcal{S}$ and $\mathcal{S}^C$,

$$e^{-\varepsilon} \le \frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} \le e^{\varepsilon}, \qquad (57)$$

$$e^{-\varepsilon} \le \frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]} \le e^{\varepsilon}. \qquad (58)$$

Thus, to achieve the left inequality in the second expression in (10),

$$\frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1]} \ge e^{-\varepsilon} \cdot (e^{\varepsilon})^{-1} = e^{-2\varepsilon}. \qquad (59)$$

To achieve the right inequality in the second expression in (10),

$$\frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1]} \le e^{\varepsilon} \cdot \left(e^{-\varepsilon}\right)^{-1} = e^{2\varepsilon}. \qquad (60)$$

□

Theorem 1. *Under Assumption 1 and Assumption 2, if the release of $T^* = t^*$ satisfies DP, then*

$$r_i(p_i, q_i, t^*) \leq \frac{1}{q_i p_i + e^{-2\varepsilon}(1-q_i)p_i + e^{-\varepsilon}(1-p_i)}. \tag{11}$$

Proof. We begin by applying Bayes' Theorem to reverse the conditional in the relative risk.

$$\begin{aligned} r_i(p_i, q_i, t^*) &= \frac{P_{\mathcal{M}}[Y_i \in \mathcal{S}, I_i = 1 \mid T^* = t^*]}{P_{\mathcal{M}}[Y_i \in \mathcal{S}, I_i = 1]} & (61) \\ &= \frac{\frac{P_{\mathcal{M}}[T^*=t^* \mid Y_i \in \mathcal{S}, I_i=1]\, P_{\mathcal{M}}[Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^*=t^*]}}{P_{\mathcal{M}}[Y_i \in \mathcal{S}, I_i = 1]} & (62) \\ &= \frac{P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]}{P_{\mathcal{M}}[T^* = t^*]}. & (63) \end{aligned}$$

We may decompose the denominator via the law of total probability.

$$\begin{aligned} P_{\mathcal{M}}[T^* = t^*] = {} & P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]\, P_{\mathcal{M}}[Y_i \in \mathcal{S} \mid I_i = 1]\, P_{\mathcal{M}}[I_i = 1] \\ & + P_{\mathcal{M}}[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1]\, P_{\mathcal{M}}[Y_i \notin \mathcal{S} \mid I_i = 1]\, P_{\mathcal{M}}[I_i = 1] \\ & + P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]\, P_{\mathcal{M}}[I_i = 0] & (64) \\ = {} & P_{\mathcal{M}}[T^* = t^* \mid Y_i \in \mathcal{S}, I_i = 1]\, q_i p_i \\ & + P_{\mathcal{M}}[T^* = t^* \mid Y_i \notin \mathcal{S}, I_i = 1]\, (1 - q_i) p_i \\ & + P_{\mathcal{M}}[T^* = t^* \mid I_i = 0]\, (1 - p_i). & (65) \end{aligned}$$

Using this expansion in the expression for r_i and dividing through by the numerator yields

$$r_i(p_i, q_i, t^*) = \frac{1}{q_i p_i + \frac{P_{\mathcal{M}}[T^*=t^* \mid Y_i \notin \mathcal{S}, I_i=1]}{P_{\mathcal{M}}[T^*=t^* \mid Y_i \in \mathcal{S}, I_i=1]}\,(1-q_i)p_i + \frac{P_{\mathcal{M}}[T^*=t^* \mid I_i=0]}{P_{\mathcal{M}}[T^*=t^* \mid Y_i \in \mathcal{S}, I_i=1]}\,(1-p_i)}. \tag{66}$$

Using Lemma 1, we then have

$$r_i(p_i, q_i, t^*) \leq \frac{1}{q_i p_i + e^{-2\varepsilon}(1-q_i)p_i + e^{-\varepsilon}(1-p_i)}. \tag{67}$$

□

Theorem 2. *For any individual i, fix the prior probabilities, p_i and q_i, and a desired bound on the relative disclosure risk, $r^*(p_i, q_i)$. Define $\varepsilon_i(p_i, q_i)$ to be the function.*

$$\varepsilon_i(p_i, q_i) = \begin{cases} \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)}-p_iq_i\right)}-(1-p_i)}\right), & \textit{if } 0 < q_i < 1; \\ \log\left(\frac{1-p_i}{\frac{1}{r^*(p_i,1)}-p_i}\right), & \textit{if } q_i = 1. \end{cases} \tag{12}$$

Under the conditions of Theorem 1, any statistic $T^* = t^*$ released under ε-DP with $\varepsilon \le \varepsilon_i(p_i, q_i)$ will satisfy $r_i(p_i, q_i, t^*) \le r^*(p_i, q_i)$.

Proof. We begin with the simpler case where $q_i = 1$. By Theorem 1,

$$r_i(p_i, 1, t^*) \le \frac{1}{p_i + e^{-\varepsilon}(1-p_i)}. \tag{68}$$

Since $\varepsilon \le \log\left((1-p_i)/(1/r^*(p_i,1) - p_i)\right)$, it follows that $e^{-\varepsilon} \ge (1/r^*(p_i,1) - p_i)/(1-p_i)$. Thus, as desired,

$$r_i(p_i, 1, t^*) \le \frac{1}{p_i + \frac{\frac{1}{r^*(p_i,1)}-p_i}{1-p_i}(1-p_i)} = \frac{1}{p_i + \left(\frac{1}{r^*(p_i,1)} - p_i\right)} = r^*(p_i, 1). \tag{69}$$

Now consider the case where $0 < q_i < 1$. By Theorem 1,

$$r_i(p_i, q_i, t^*) \le \frac{1}{q_ip_i + e^{-2\varepsilon}(1-q_i)p_i + e^{-\varepsilon}(1-p_i)}. \tag{70}$$

Since

$$\varepsilon \le \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right)} - (1-p_i)}\right), \tag{71}$$

it follows that

$$e^{-\varepsilon} \ge \frac{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right)} - (1-p_i)}{2p_i(1-q_i)}. \tag{72}$$

Taking the square gives

$$e^{-2\varepsilon} \ge \frac{4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right) + 2(1-p_i)^2}{4p_i^2(1-q_i)^2} - \frac{2(1-p_i)\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right)}}{4p_i^2(1-q_i)^2} \tag{73}$$

$$= \frac{\frac{1}{r^*(p_i,q_i)} - p_iq_i}{p_i(1-q_i)} + \frac{(1-p_i)^2 - (1-p_i)\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_iq_i\right)}}{2p_i^2(1-q_i)^2}. \tag{74}$$

Thus,

$$\begin{aligned}
q_i p_i + {} & e^{-2\varepsilon}(1-q_i)p_i + e^{-\varepsilon}(1-p_i) \\
& \geq q_i p_i + \frac{\frac{1}{r^*(p_i,q_i)} - p_i q_i}{p_i(1-q_i)} \cdot (1-q_i)p_i \\
& \quad + \frac{(1-p_i)^2 - (1-p_i)\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_i q_i\right)}}{2p_i^2(1-q_i)^2} \cdot (1-q_i)p_i \\
& \quad + \frac{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_i q_i\right)} - (1-p_i)}{2p_i(1-q_i)} \cdot (1-p_i) \qquad (75) \\
& = q_i p_i + \left(\frac{1}{r^*(p_i,q_i)} - p_i q_i\right) \\
& \quad + \left((1-p_i) - \sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_i q_i\right)}\right)\frac{1-p_i}{2p_i(1-q_i)} \\
& \quad - \left((1-p_i) - \sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_i q_i\right)}\right)\frac{1-p_i}{2p_i(1-q_i)} \qquad (76) \\
& = \frac{1}{r^*(p_i,q_i)}. \qquad (77)
\end{aligned}$$

It is then immediate that

$$r_i(p_i,q_i,t^*) \leq \frac{1}{q_i p_i + e^{-2\varepsilon}(1-q_i)p_i + e^{-\varepsilon}(1-p_i)} \leq \frac{1}{\frac{1}{r^*(p_i,q_i)}} = r^*(p_i,q_i). \qquad (78)$$

□

Corollary 1. *Under the conditions of Theorem 1, for all $p_i, q_i \in (0, 1]$ and all t^*,*

$$r_i(p_i, q_i, t^*) \leq e^{2\varepsilon}. \tag{79}$$

Proof. First note that since, $0 \leq e^{-\varepsilon} \leq 1$, it follows that $e^{-2\varepsilon} \leq e^{-\varepsilon}$. Then, applying the result of Theorem 1,

$$r_i(p_i, q_i, t^*) \leq \frac{1}{q_i p_i + e^{-2\varepsilon}(1-q_i)p_i + e^{-\varepsilon}(1-p_i)} \leq \frac{1}{q_i p_i + e^{-2\varepsilon}(1-q_i)p_i + e^{-2\varepsilon}(1-p_i)} \tag{80}$$

Combining terms and using the fact that $q_i p_i \geq 0$ gives

$$r_i(p_i, q_i, t^*) \leq \frac{1}{q_i p_i + e^{-2\varepsilon}(1-q_i p_i)} \leq \frac{1}{e^{-2\varepsilon} + q_i p_i(1-e^{-2\varepsilon})} \leq \frac{1}{e^{-2\varepsilon}+0} = e^{2\varepsilon}. \tag{81}$$

□

Corollary 2. *Under the conditions of Theorem 2, if $p_i = 1$ and $0 < q_i < 1$, then any statistic $T^* = t^*$ released under ε-DP with*

$$\varepsilon \leq \frac{1}{2}\log\left(\frac{1-q_i}{\frac{1}{r^*(1,q_i)} - q_i}\right), \tag{82}$$

will satisfy $r_i(1, q_i, t^*) \leq r^*(1, q_i)$.

Proof. Plugging $p_i = 1$ into the expression from Theorem 2 yields

$$\begin{aligned}
\varepsilon &\leq \log\left(\frac{2(1-q_i)}{\sqrt{0 + 4(1-q_i)\left(\frac{1}{r^*(1,q_i)} - q_i\right)} - 0}\right) && (83)\\
&= \log\left(\sqrt{\frac{1-q_i}{\frac{1}{r^*(1,q_i)} - q_i}}\right) && (84)\\
&= \frac{1}{2}\log\left(\frac{1-q_i}{\frac{1}{r^*(1,q_i)} - q_i}\right). && (85)
\end{aligned}$$

□

Lemma 2. *Fix $p_i, q_i \in (0,1)$ and let $r^*(p_i,q_i) = \tilde{a}/(p_i q_i)$. Then the function*

$$\varepsilon(p_i,q_i) = \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)}-p_iq_i\right)}-(1-p_i)}\right) \tag{86}$$

has partial derivatives such that

1. $\frac{\partial\varepsilon(p_i,q_i)}{\partial p_i} < 0$ *for all* $0 < p_i < 1$ *and* $0 < q_i < 1$
2. $\frac{\partial\varepsilon(p_i,q_i)}{\partial q_i} < 0$ *for all* $0 < p_i < 1$ *and* $0 < q_i < 1$.

Proof. To begin, we re-express the function of interest in the form

$$\begin{aligned}\varepsilon(p_i,q_i) &= \log\left(2p_i(1-q_i)\right) - \log\left(\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{p_iq_i}{\tilde{a}}-p_iq_i\right)}-(1-p_i)\right) &(87)\\ &= \log\left(2p_i(1-q_i)\right) - \log\left(\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}-(1-p_i)\right). &(88)\end{aligned}$$

We first examine $\partial\varepsilon(p_i,q_i)/\partial p_i$. Taking the partial derivative with respect to q_i gives

$$\begin{aligned}\frac{\partial\varepsilon(p_i,q_i)}{\partial p_i} &= \frac{1}{p_i} - \frac{\frac{-2(1-p_i)+8p_iq_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}{2\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}}+1}{\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}-(1-p_i)} &(89)\\ &= \frac{1}{p_i} + \frac{\frac{(1-p_i)-4p_iq_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)-\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}}{\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}}}{\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}-(1-p_i)} &(90)\\ &= \frac{\frac{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)+p_i(1-p_i)-4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)-[(1-p_i)-p_i]\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}}{\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}}}{p_i\left[\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}-(1-p_i)\right]} &(91)\\ &= \frac{\frac{[(1-p_i)+p_i](1-p_i)-\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}}{\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}}}{p_i\left[\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}-(1-p_i)\right]} &(92)\\ &= \frac{\frac{-\left[\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}-(1-p_i)\right]}{\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}}}{p_i\left[\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}-(1-p_i)\right]} &(93)\\ &= -\frac{1}{p_i\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}}. &(94)\end{aligned}$$

Certainly, the denominator of (94) is positive. Thus, $\partial\varepsilon(p_i,q_i)/\partial p_i < 0$, as desired.

We now examine $\partial\varepsilon(p_i, q_i)/\partial q_i$. Taking the partial derivative with respect to q_i gives

$$\frac{\partial\varepsilon(p_i,q_i)}{\partial q_i} = -\frac{1}{1-q_i} - \frac{\frac{4p_i^2\left(\frac{1}{\tilde{a}}-1\right)\frac{\partial}{\partial q_i}[q_i(1-q_i)]}{2\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}}}{\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}-(1-p_i)} \tag{95}$$

$$= -\frac{\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}-(1-p_i)+\frac{2p_i^2\left(\frac{1}{\tilde{a}}-1\right)(1-q_i)\frac{\partial}{\partial q_i}[q_i(1-q_i)]}{\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}}}{(1-q_i)\left[\sqrt{(1-p_i)^2+4^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}-(1-p_i)\right]} \tag{96}$$

$$= -\frac{\frac{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)-(1-p_i)\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}+2p_i^2\left(\frac{1}{\tilde{a}}-1\right)(1-q_i)\frac{\partial}{\partial q_i}[q_i(1-q_i)]}{\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}}}{(1-q_i)\left[\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}-(1-p_i)\right]}. \tag{97}$$

Consider the three parts of (97). Since $\tilde{a} < 1$, we have that in the denominator of the numerator, $\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)} > 0$. The denominator is also positive, since $(1-q_i) > 0$ and

$$\left[\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)}-(1-p_i)\right] > \left[\sqrt{(1-p_i)^2}-(1-p_i)\right] = 0. \tag{98}$$

This leaves the numerator of the numerator of (97). Let us denote this expression as $g(q_i)$. If $g(q_i) > 0$, then it follows that $\partial\varepsilon(p_i, q_i)/\partial q_i < 0$. To show this, we begin by simplifying the expression for $g(q_i)$:

$$\begin{aligned} g(q_i) &= (1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)-(1-p_i)\sqrt{(1-p_i)^2+4p_i^2q_i(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)} \\ &\quad + 2p_i^2\left(\frac{1}{\tilde{a}}-1\right)(1-q_i)\frac{\partial}{\partial q_i}[q_i(1-q_i)] \qquad (99) \\ &= (1-p_i)^2+2p_i^2(1-q_i)(2q_i)\left(\frac{1}{\tilde{a}}-1\right)-(1-p_i)\sqrt{(1-p_i)^2+4p_i^2(1-q_i)q_i\left(\frac{1}{\tilde{a}}-1\right)} \\ &\quad + 2p_i^2(1-q_i)\left(\frac{1}{\tilde{a}}-1\right)(1-2q_i) \qquad (100) \\ &= (1-p_i)^2-(1-p_i)\sqrt{(1-p_i)^2+4p_i^2(1-q_i)q_i\left(\frac{1}{\tilde{a}}-1\right)}+2p_i^2(1-q_i)\left(\frac{1}{\tilde{a}}-1\right). \qquad (101) \end{aligned}$$

Taking the derivative of g, we find that

$$\begin{aligned} \frac{\partial g(q_i)}{\partial q_i} &= -(1-p_i)\frac{4p_i^2(1-2q_i)\left(\frac{1}{\tilde{a}}-1\right)}{2\sqrt{(1-p_i)^2+4p_i^2(1-q_i)q_i\left(\frac{1}{\tilde{a}}-1\right)}}-2p_i^2\left(\frac{1}{\tilde{a}}-1\right) \qquad (102) \\ &= -2p_i^2\left(\frac{1}{\tilde{a}}-1\right)\left[(1-p_i)\frac{1-2q_i}{\sqrt{(1-p_i)^2+4p_i^2(1-q_i)q_i\left(\frac{1}{\tilde{a}}-1\right)}}+1\right]. \qquad (103) \end{aligned}$$

For $0 < q_i \leq 1/2$, $1 - 2q_i \geq 0$ and so $\partial g(q_i)/\partial q_i < 0$. For $1/2 < q_i < 1$, since $2q_i - 1 < 1$ and $\sqrt{(1-p_i)^2 + 4p_i^2(1-q_i)q_i\,(1/\tilde{a} - 1)} > 1 - p_i$,

$$\frac{\partial g(q_i)}{\partial q_i} = -2p_i^2\left(\frac{1}{\tilde{a}} - 1\right)\left[1 - (1-p_i)\,\frac{2q_i - 1}{\sqrt{(1-p_i)^2 + 4p_i^2(1-q_i)q_i\left(\frac{1}{\tilde{a}} - 1\right)}}\right] \tag{104}$$

$$< -2p_i^2\left(\frac{1}{\tilde{a}} - 1\right)\left[1 - (1-p_i)\,\frac{1}{1-p_i}\right] \tag{105}$$

$$< 0. \tag{106}$$

Thus, $g(q_i)$ strictly decreasing as a function of q_i on $0 < q_i < 1$. Since

$$g\,(1) = (1-p_i)^2 - (1-p_i)\sqrt{(1-p_i)^2 + 4p_i^2(1-1)1\left(\frac{1}{\tilde{a}} - 1\right)} + 2p_i^2(1-1)\left(\frac{1}{\tilde{a}} - 1\right) = 0, \tag{107}$$

it follows that $g(q_i) > 0$ for $0 < q_i < 1$ and so $\partial\varepsilon(p_i, q_i)/\partial q_i < 0$ in this range. □

Lemma 3. *Fix $p_i, q_i \in (0,1)$ and let $r^*(p_i, q_i) = \tilde{r} < 1/(p_i q_i)$. Then the function*

$$\varepsilon(p_i, q_i) = \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{r^*(p_i,q_i)} - p_i q_i\right)} - (1-p_i)}\right) \tag{108}$$

has partial derivatives such that

1. $\frac{\partial\varepsilon(p_i,q_i)}{\partial p_i} < 0$ *if* $0 < q_i < \frac{1}{\tilde{r}+1}$ *and* $0 < p_i < 1$
2. $\frac{\partial\varepsilon(p_i,q_i)}{\partial p_i} = 0$ *if* $q_i = \frac{1}{\tilde{r}+1}$ *and* $0 < p_i < 1$
3. $\frac{\partial\varepsilon(p_i,q_i)}{\partial p_i} > 0$ *if* $\frac{1}{\tilde{r}+1} < q_i < 1$ *and* $0 < p_i < \frac{1}{q_i\tilde{r}}$
4. $\frac{\partial\varepsilon(p_i,q_i)}{\partial q_i} > 0$ *if* $0 < p_i < 1$ *and* $0 < q_i < \frac{1}{p_i\tilde{r}}$.

Proof. To begin, we re-express the function of interest in the form

$$\varepsilon(p_i, q_i) = \log\left(2p_i(1-q_i)\right) - \log\left(\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_i q_i\right)} - (1-p_i)\right). \tag{109}$$

We now examine $\partial\varepsilon(p_i,q_i)/\partial p_i$. Taking the partial derivative of with respect to p_i and simplifying gives

$$\frac{\partial\varepsilon(p_i,q_i)}{\partial p_i} = \frac{1}{p_i} - \frac{\frac{\frac{\partial}{\partial p_i}\left[(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)\right]}{2\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}} + 1}{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_iq_i\right)} - (1-p_i)} \tag{110}$$

$$= \frac{1}{p_i} - \frac{\frac{-2(1-p_i)+4(1-q_i)(\frac{1}{\tilde{r}}-p_iq_i)-4p_i(1-q_i)q_i+2\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}{2\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}}{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_iq_i\right)} - (1-p_i)} \tag{111}$$

$$= \frac{\frac{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)-(1-p_i)\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}}{p_i\left[\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_iq_i\right)} - (1-p_i)\right]}$$

$$+ \frac{\frac{p_i(1-p_i)-2p_i(1-q_i)(\frac{1}{\tilde{r}}-p_iq_i)+2p_i^2(1-q_i)q_i-p_i\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}}{p_i\left[\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_iq_i\right)} - (1-p_i)\right]} \tag{112}$$

$$= \frac{\frac{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)+p_i(1-p_i)-2p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)+2p_i^2(1-q_i)q_i-\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}}{p_i\left[\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_iq_i\right)} - (1-p_i)\right]} \tag{113}$$

$$= \frac{2p_i(1-q_i)\frac{1}{\tilde{r}} + (1-p_i) - \sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_iq_i\right)}}{p_i\left[\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_iq_i\right)} - (1-p_i)\right]\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_iq_i\right)}}. \tag{114}$$

We can rearrange the expression in (114) to the following form.

$$\frac{\partial\varepsilon(p_i,q_i)}{\partial p_i} = \frac{\frac{2p_i(1-q_i)\frac{1}{\tilde{r}}}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}-(1-p_i)} - 1}{p_i\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_iq_i\right)}}. \tag{115}$$

The denominator of (115) is certainly always positive, so the sign of $\partial\varepsilon(p_i, q_i)/\partial p_i$ is determined by the sign of the numerator. To determine the sign, we will use the following equality.

$$\begin{aligned}
&2p_i(1-q_i)\frac{1}{\tilde{r}} \\
&\quad = \sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_i q_i\right) + 4p_i^2(1-q_i)\frac{\tilde{r}^2-1}{\tilde{r}^2}\left(q_i - \frac{1}{\tilde{r}+1}\right)} - (1-p_i).
\end{aligned} \tag{116}$$

To demonstrate that (116) holds, note the following.

$$\begin{align}
\left(2p_i(1-q_i)\frac{1}{\tilde{r}} + (1-p_i)\right)^2 &= 4p_i^2(1-q_i)^2\frac{1}{\tilde{r}^2} + 4p_i(1-p_i)(1-q_i)\frac{1}{\tilde{r}} + (1-p_i)^2 \tag{117}\\
&= 4p_i^2(1-q_i)^2\frac{1}{\tilde{r}^2} + 4p_i(1-p_i)(1-q_i)\frac{1}{\tilde{r}} - 4p_i^2(1-q_i)\frac{\tilde{r}^2-1}{\tilde{r}^2}\left(q_i - \frac{1}{\tilde{r}+1}\right) \notag\\
&\qquad + 4p_i^2(1-q_i)\frac{\tilde{r}^2-1}{\tilde{r}^2}\left(q_i - \frac{1}{\tilde{r}+1}\right) + (1-p_i)^2 \tag{118}\\
&= 4p_i(1-q_i)\left[p_i(1-q_i)\frac{1}{\tilde{r}^2} + (1-p_i)\frac{1}{\tilde{r}} - p_i\frac{\tilde{r}^2-1}{\tilde{r}^2}\left(q_i - \frac{1}{\tilde{r}+1}\right)\right] \notag\\
&\qquad + 4p_i^2(1-q_i)\frac{\tilde{r}^2-1}{\tilde{r}^2}\left(q_i - \frac{1}{\tilde{r}+1}\right) + (1-p_i)^2 \tag{119}\\
&= 4p_i(1-q_i)\left[\frac{1}{\tilde{r}} + p_i\left(\frac{1}{\tilde{r}^2} - \frac{1}{\tilde{r}} + \frac{\tilde{r}-1}{\tilde{r}^2}\right) - p_i q_i\left(\frac{1}{\tilde{r}^2} + \frac{\tilde{r}^2-1}{\tilde{r}^2}\right)\right] \notag\\
&\qquad + 4p_i^2(1-q_i)\frac{\tilde{r}^2-1}{\tilde{r}^2}\left(q_i - \frac{1}{\tilde{r}+1}\right) + (1-p_i)^2 \tag{120}\\
&= 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_i q_i\right) + 4p_i^2(1-q_i)\frac{\tilde{r}^2-1}{\tilde{r}^2}\left(q_i - \frac{1}{\tilde{r}+1}\right) + (1-p_i)^2. \tag{121}
\end{align}$$

Rearranging (121) gives (116).

Now note that, from (116), if $q_i = 1/(\tilde{r}+1)$, then

$$2p_i(1-q_i)\frac{1}{\tilde{r}} = \sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_i q_i\right)} - (1-p_i). \tag{122}$$

It follows that the ratio in the numerator of (115) equals 1 and thus $\partial\varepsilon(p_i, q_i)/\partial p_i = 0$. If $0 < q_i < \frac{1}{\tilde{r}+1}$, then the term $4p_i^2(1-q_i)(\tilde{r}^2-1)/\tilde{r}^2(q_i - 1/(\tilde{r}+1)) < 0$ and so

$$2p_i(1-q_i)\frac{1}{\tilde{r}} < \sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_i q_i\right)} - (1-p_i). \tag{123}$$

It follows that the ratio in the numerator of (115) is less than 1 and thus $\partial\varepsilon(p_i, q_i)/\partial p_i < 0$. If $\frac{1}{\tilde{r}+1} < q_i < 1$, then the term $4p_i^2(1-q_i)(\tilde{r}^2-1)/\tilde{r}^2(q_i - 1/(\tilde{r}+1)) > 0$ and so

$$2p_i(1-q_i)\frac{1}{\tilde{r}} > \sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_i q_i\right)} - (1-p_i). \tag{124}$$

It follows that the ratio in the numerator of (115) is greater than 1 and thus $\partial\varepsilon(p_i, q_i)/\partial p_i > 0$.

We now examine $\partial\varepsilon(p_i,q_i)/\partial q_i$. Taking the partial derivative of with respect to q_i gives

$$\frac{\partial\varepsilon(p_i,q_i)}{\partial q_i} = -\frac{1}{1-q_i} - \frac{\frac{4p_i\frac{\partial}{\partial q_i}\left[(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)\right]}{2\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}-(1-p_i)} \tag{125}$$

$$= -\frac{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}-(1-p_i)+\frac{2p_i(1-q_i)\frac{\partial}{\partial q_i}\left[(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)\right]}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}}{(1-q_i)\left[\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}-(1-p_i)\right]} \tag{126}$$

$$= -\frac{\frac{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)-(1-p_i)\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}+2p_i(1-q_i)\frac{\partial}{\partial q_i}\left[(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)\right]}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}}{(1-q_i)\left[\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}-(1-p_i)\right]}. \tag{127}$$

Consider the three parts of (127). Since $\tilde{r} < 1/(p_iq_i)$, it follows that $(1/\tilde{r} - p_iq_i) > 0$. Thus, in the denominator of the numerator, $\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(1/\tilde{r}-p_iq_i\right)} > 0$. The denominator is also positive, since $(1-q_i) > 0$ and

$$\left[\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}-(1-p_i)\right] > \left[\sqrt{(1-p_i)^2}-(1-p_i)\right] = 0. \tag{128}$$

This leaves the numerator of the numerator of (127). Let us denote this expression as $g(q_i)$. If $g(q_i) < 0$, then it follows that $\partial\varepsilon(p_i,q_i)/\partial q_i > 0$. To show this, we begin by simplifying the expression for $g(q_i)$:

$$g(q_i) = (1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)-(1-p_i)\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}$$
$$+2p_i(1-q_i)\frac{\partial}{\partial q_i}\left[(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)\right] \tag{129}$$

$$= (1-p_i)^2+2p_i(1-q_i)\left(\frac{2}{\tilde{r}}-2p_iq_i\right)-(1-p_i)\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}$$
$$+2p_i(1-q_i)\left(2p_iq_i-p_i-\frac{1}{\tilde{r}}\right) \tag{130}$$

$$= (1-p_i)^2-(1-p_i)\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}+2p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_i\right). \tag{131}$$

Taking the derivative of g, we find that

$$\frac{\partial g(q_i)}{\partial q_i} = -(1-p_i)\frac{4p_i\left(2p_iq_i-p_i-\frac{1}{\tilde{r}}\right)}{2\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}-2p_i\left(\frac{1}{\tilde{r}}-p_i\right) \tag{132}$$

$$= 2p_i\left[(1-p_i)\frac{2p_i(1-q_i)+\left(\frac{1}{\tilde{r}}-p_i\right)}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}-\left(\frac{1}{\tilde{r}}-p_i\right)\right]. \tag{133}$$

Note that since $\tilde{r} < \frac{1}{p_i q_i}$, it follows that $\left(\frac{1}{\tilde{r}} - p_i q_i\right) > 0$ and $p_i(1 - q_i) > p_i - \frac{1}{r}$. Thus,

$$\begin{aligned}
\frac{\partial g(q_i)}{\partial q_i} &= 2p_i \left[(1 - p_i) \frac{2p_i(1 - q_i) - \left(p_i - \frac{1}{\tilde{r}}\right)}{\sqrt{(1 - p_i)^2 + 4p_i(1 - q_i)\left(\frac{1}{\tilde{r}} - p_i q_i\right)}} + \left(p_i - \frac{1}{\tilde{r}}\right) \right] && (134)\\
&> 2p_i \left[(1 - p_i) \frac{\left(p_i - \frac{1}{\tilde{r}}\right)}{\sqrt{(1 - p_i)^2}} + \left(p_i - \frac{1}{\tilde{r}}\right) \right] && (135)\\
&= 4p_i \left(p_i - \frac{1}{\tilde{r}}\right) && (136)\\
&> 0. && (137)
\end{aligned}$$

Thus, $g(q_i)$ is strictly increasing for $\tilde{r} < 1/(p_i q_i)$. Note that in terms of q_i, this is equivalent to the range $0 < q_i < 1/(p_i \tilde{r})$. Since

$$\begin{aligned}
&g\left(\frac{1}{p_i \tilde{r}}\right)\\
&= (1 - p_i)^2 - (1 - p_i)\sqrt{(1 - p_i)^2 + 4p_i(1 - \frac{1}{p_i \tilde{r}})\left(\frac{1}{\tilde{r}} - p_i \frac{1}{p_i \tilde{r}}\right)} + 2p_i \left(1 - \frac{1}{p_i \tilde{r}}\right)\left(\frac{1}{\tilde{r}} - p_i\right) && (138)\\
&= 2p_i \left(1 - \frac{1}{p_i \tilde{r}}\right)\left(\frac{1}{\tilde{r}} - p_i\right) && (139)\\
&< 0, && (140)
\end{aligned}$$

it follows that $g(q_i) < 0$ for $0 < q_i < 1/(p_i \tilde{r})$. Thus, as desired, $\partial \varepsilon(p_i, q_i)/\partial q_i > 0$ in this range. □

Theorem 3. *Under the conditions of Theorem 2, if $r^*(p_i, q_i) = \tilde{r} > 1$, the solution to the minimization problem in (13) is*

$$\varepsilon = \frac{1}{2}\log\left(\tilde{r}\right). \tag{141}$$

Proof. We define $\varepsilon(p_i, q_i)$ as follows.

$$\varepsilon(p_i, q_i) = \begin{cases} \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}-(1-p_i)}\right), & \text{if } 0 < q_i < 1; \\ \log\left(\frac{1-p_i}{\frac{1}{\tilde{r}}-p_i}\right), & \text{if } q_i = 1. \end{cases} \tag{142}$$

The solution to the minimization problem in (13) corresponds to the minimum of $\varepsilon(p_i, q_i)$ over $(0, 1] \times (0, 1]$. Lemma 3 implies that $\partial\varepsilon(p_i, q_i)/\partial q_i \neq 0$ for any $0 < q_i < 1$. This means that this function must take its minimum value around its boundary, i.e., when $p_i = 1$, $q_i = 1$, $p_i \to 0$, or $q_i \to 0$. We examine each of these in turn.

We begin with the boundary where $p_i = 1$. By Corollary 2, for $0 < q_i < 1$,

$$\varepsilon(1, q_i) = \frac{1}{2}\log\left(\frac{1-q_i}{\frac{1}{\tilde{r}}-q_i}\right). \tag{143}$$

We take the partial derivative of this function with respect to q_i.

$$\frac{\partial\varepsilon(1, q_i)}{\partial q_i} = \frac{1}{2}\left[-\frac{1}{1-q_i}+\frac{1}{\frac{1}{\tilde{r}}-q_i}\right]. \tag{144}$$

Since $1/\tilde{r} < 1$, it follows that $\partial\varepsilon(1, q_i)/\partial q_i > 0$ for all $0 < q_i < 1$. Thus, the minimum occurs as $q_i \to 0$ and is

$$\lim_{q_i \to 0} \varepsilon(1, q_i) = \frac{1}{2}\log(\tilde{r}). \tag{145}$$

We next examine with the boundary where $q_i = 1$. When $0 < p_i < 1$,

$$\varepsilon(p_i, 1) = \log\left(\frac{1-p_i}{\frac{1}{\tilde{r}}-p_i}\right). \tag{146}$$

We take the partial derivative of this function with respect to p_i.

$$\frac{\partial\varepsilon(p_i, 1)}{\partial p_i} = -\frac{1}{1-p_i}+\frac{1}{\frac{1}{\tilde{r}}-p_i}. \tag{147}$$

Since $1/\tilde{r} < 1$, it follows that $\partial\varepsilon(p_i, 1)/\partial p_i > 0$ for all $0 < p_i < 1$. Thus, the minimum occurs as $p_i \to 0$ and is

$$\lim_{p_i \to 0} \varepsilon(p_i, 1) = \log(\tilde{r}). \tag{148}$$

We next examine the boundary where $p_i \to 0$. Since the logarithm is a continuous function, for all $0 < q_i < 1$,

$$\lim_{p_i \to 0} \varepsilon(p_i, q_i) = \lim_{p_i \to 0} \log\left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}-(1-p_i)}\right) \tag{149}$$

$$= \log\left(\lim_{p_i \to 0} \frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}-(1-p_i)}\right). \tag{150}$$

Using L'Hôpital's Rule,

$$\lim_{p_i \to 0} \varepsilon(p_i, q_i) = \log\left(\lim_{p_i \to 0} \frac{2(1-q_i)}{\frac{-2(1-p_i)+4(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)-4p_iq_i(1-q_i)}{2\sqrt{(1-p_i)^2+4p_i(1-q_i)\left(\frac{1}{\tilde{r}}-p_iq_i\right)}}+1}\right) \tag{151}$$

$$= \log\left(\frac{2(1-q_i)}{\frac{-1+2(1-q_i)\left(\frac{1}{\tilde{r}}\right)}{\sqrt{1}}+1}\right) = \log(\tilde{r}). \tag{152}$$

Finally, we examine the boundary where $q_i \to 0$. For all $0 < p_i \leq 1$

$$\lim_{q_i \to 0} \varepsilon(p_i, q_i) = \lim_{q_i \to 0} \log \left(\frac{2p_i(1-q_i)}{\sqrt{(1-p_i)^2 + 4p_i(1-q_i)\left(\frac{1}{\tilde{r}} - p_i q_i\right)} - (1-p_i)} \right) \tag{153}$$

$$= \log \left(\frac{2p_i}{\sqrt{(1-p_i)^2 + 4p_i\left(\frac{1}{\tilde{r}}\right)} - (1-p_i)} \right). \tag{154}$$

We take the partial derivative with respect to p_i. From (115) in the proof of Lemma 3, we have that

$$\frac{\partial}{\partial p_i} \left(\lim_{q_i \to 0} \varepsilon(p_i, q_i) \right) = \frac{\partial}{\partial p_i} \left(\log \left(\frac{2p_i}{\sqrt{(1-p_i)^2 + 4p_i\left(\frac{1}{\tilde{r}} - p_i\right)} - (1-p_i)} \right) \right) \tag{155}$$

$$= \frac{\frac{2p_i \frac{1}{\tilde{r}}}{\sqrt{(1-p_i)^2 + 4p_i\left(\frac{1}{\tilde{r}}\right)} - (1-p_i)} - 1}{p_i \sqrt{(1-p_i)^2 + 4p_i\left(\frac{1}{\tilde{r}}\right)}}. \tag{156}$$

By (116) with $q_i = 0$, the ratio in the numerator of (156) is less than 1 and so $\partial / \partial p_i \left(\lim_{q_i \to 0} \varepsilon(p_i, q_i) \right)$ is negative for all $0 < p_i \leq 1$. Thus, the minimum occurs at $p_i = 1$ and is as in (145). The global minimum of $\varepsilon(p_i, q_i)$ is thus $1/2 \log(\tilde{r})$. □

Theorem 4. *Under the conditions of Theorem 2, let $\tilde{a} < 1$, $\tilde{p} \leq 1$, and $\tilde{r} > 1$, and $0 < q_i \leq 1$. If the function r^* is such that $r^*(\tilde{p}, q_i) = \max\{\tilde{a}/(\tilde{p}q_i), \tilde{r}\}$ and $r^*(p_i, q_i) = \infty$ if $p_i \neq \tilde{p}$, then the solution to the minimization problem in (13) is*

$$\varepsilon = \begin{cases} \log\left(\frac{\tilde{a}(1-\tilde{p})}{\tilde{p}(1-\tilde{a})}\right), & \textit{if } 0 < \tilde{p} \leq \frac{\tilde{a}}{\tilde{r}}; \\ \log\left(\frac{2(\tilde{p}\tilde{r}-\tilde{a})}{\sqrt{\tilde{r}^2(1-\tilde{p})^2+4(\tilde{p}\tilde{r}-\tilde{a})(1-\tilde{a})}-\tilde{r}(1-\tilde{p})}\right), & \textit{if } \frac{\tilde{a}}{\tilde{r}} < \tilde{p} \leq 1. \end{cases} \tag{157}$$

Proof. We define $\varepsilon(q_i)$ as follows.

$$\varepsilon(q_i) = \begin{cases} \log\left(\frac{2\tilde{p}(1-q_i)}{\sqrt{(1-\tilde{p})^2+4\tilde{p}(1-q_i)\left(\frac{1}{r^*(\tilde{p},q_i)}-\tilde{p}q_i\right)}-(1-\tilde{p})}\right), & \text{if } 0 < q_i < 1; \\ \log\left(\frac{1-\tilde{p}}{\frac{1}{r^*(\tilde{p},q_i)}-\tilde{p}}\right), & \text{if } q_i = 1. \end{cases} \tag{158}$$

The solution to the minimization problem in (13) corresponds to the minimum of $\varepsilon(q_i)$ over $(0, 1]$.

We begin with the case where $0 < \tilde{p} \leq \tilde{a}/\tilde{r}$. In this setting, we have that $\tilde{r} \leq \tilde{a}/\tilde{p} \leq \tilde{a}/(\tilde{p}q_i)$, so $r^*(\tilde{p}, q_i) = \tilde{a}/(\tilde{p}q_i)$ for all q_i. By Lemma 2, $\varepsilon(q_i)$ is a decreasing function of q_i, so the minimum occurs at $q_i = 1$ and is

$$\varepsilon(1) = \log\left(\frac{1-\tilde{p}}{\frac{1}{r^*(\tilde{p},1)}-\tilde{p}}\right) = \log\left(\frac{1-\tilde{p}}{\frac{\tilde{p}}{\tilde{a}}-\tilde{p}}\right) = \log\left(\frac{\tilde{a}(1-\tilde{p})}{\tilde{p}(1-\tilde{a})}\right). \tag{159}$$

We now consider the case where $\tilde{a}/\tilde{r} < \tilde{p} \leq 1$. Note that $\tilde{r} \geq \tilde{a}/(\tilde{p}q_i)$ whenever $q_i \geq \tilde{a}/(\tilde{p}\tilde{r})$. Since in this setting $\tilde{a}/(\tilde{p}\tilde{r}) < 1$, the risk bound can be written in piece-wise form

$$r^*(\tilde{p}, q_i) = \begin{cases} \frac{\tilde{a}}{\tilde{p}q_i}, & \text{if } 0 < q_i < \frac{\tilde{a}}{\tilde{p}\tilde{r}}; \\ \tilde{r}, & \text{if } \frac{\tilde{a}}{\tilde{p}\tilde{r}} \leq q_i < 1. \end{cases} \tag{160}$$

Note that if $q_i \geq 1/(\tilde{r}\tilde{p})$, then $\tilde{r} \geq 1/(\tilde{p}q_i)$. Thus, $r^*(\tilde{p}, q_i) \geq \tilde{r} \geq 1/(\tilde{p}q_i)$ and so

$$\begin{aligned} &\sqrt{(1-\tilde{p})^2 + 4\tilde{p}(1-q_i)\left(\frac{1}{r^*(\tilde{p},q_i)} - \tilde{p}q_i\right)} - (1-\tilde{p}) \\ &\qquad \leq \sqrt{(1-\tilde{p})^2 + 4\tilde{p}(1-q_i)\left(\tilde{p}q_i - \tilde{p}q_i\right)} - (1-\tilde{p}) \qquad (161) \\ &\qquad = \sqrt{(1-\tilde{p})^2} - (1-\tilde{p}) \qquad (162) \\ &\qquad = 0. \qquad (163) \end{aligned}$$

Thus, $\varepsilon(q_i) = \infty$ for any $q_i \geq 1/(\tilde{r}\tilde{p})$. We may thus restrict our attention to $q_i < \min\{1, 1/(\tilde{r}\tilde{p})\}$.

Since $r^*(\tilde{p}, q_i)$ is a continuous function of q_i and $\varepsilon(q_i)$ is a continuous function of $r^*(\tilde{p}, q_i)$, it follows that $\varepsilon(q_i)$ is a continuous function of q_i. By Lemma 2, $\varepsilon(q_i)$ is a decreasing function of q_i for $0 < q_i < \tilde{a}/(\tilde{p}\tilde{r})$ and by Lemma 3, $\varepsilon(q_i)$ is an increasing function of q_i for $\tilde{a}/(\tilde{p}\tilde{r}) < q_i < \min\{1, 1/(\tilde{r}\tilde{p})\}$. Thus, the minimum must occur at $q_i = \tilde{a}/(\tilde{p}\tilde{r})$ and is

$$\begin{aligned} \varepsilon\left(\frac{\tilde{a}}{\tilde{p}\tilde{r}}\right) &= \log\left(\frac{2\tilde{p}\left(1-\frac{\tilde{a}}{\tilde{p}\tilde{r}}\right)}{\sqrt{(1-\tilde{p})^2 + 4\tilde{p}\left(1-\frac{\tilde{a}}{\tilde{p}\tilde{r}}\right)\left(\frac{1}{\tilde{r}} - \tilde{p}\frac{\tilde{a}}{\tilde{p}\tilde{r}}\right)} - (1-\tilde{p})}\right) \qquad (164) \\ &= \log\left(\frac{2\left(\tilde{p}\tilde{r}-\tilde{a}\right)}{\sqrt{\tilde{r}^2(1-\tilde{p})^2 + 4\left(\tilde{p}\tilde{r}-\tilde{a}\right)\left(1-\tilde{a}\right)} - \tilde{r}(1-\tilde{p})}\right). \qquad (165) \end{aligned}$$

□

Theorem 5. *Under the conditions of Theorem 2, let $\tilde{a} < 1$, $\tilde{q} \leq 1$, and $\tilde{r} > 1$, and $0 < p_i < 1$. If the function r^* is such that $r^*(p_i, \tilde{q}) = \max\{\tilde{a}/(p_i\tilde{q}), \tilde{r}\}$ and $r^*(p_i, q_i) = \infty$ for $q_i \neq \tilde{q}$, then the solution to the minimization problem in (13) is*

$$\varepsilon = \begin{cases} \frac{1}{2}\log\left(\frac{\tilde{a}(1-\tilde{q})}{\tilde{q}(1-\tilde{a})}\right), & \textit{if } 0 < \tilde{q} \leq \frac{\tilde{a}}{\tilde{r}}; \\ \frac{1}{2}\log\left(\frac{1-\tilde{q}}{\frac{1}{\tilde{r}}-\tilde{q}}\right), & \textit{if } 0 < \tilde{q} \leq \frac{1}{\tilde{r}+1} \textit{ and } \frac{\tilde{a}}{\tilde{r}} < \tilde{q} < 1; \\ \log\left(\frac{2\tilde{a}(1-\tilde{q})}{\sqrt{(\tilde{r}\tilde{q}-\tilde{a})^2+4\tilde{a}\tilde{q}(1-\tilde{q})(1-\tilde{a})}-(\tilde{r}\tilde{q}-\tilde{a})}\right), & \textit{if } \frac{1}{\tilde{r}+1} < \tilde{q} < 1 \textit{ and } \frac{\tilde{a}}{\tilde{r}} < \tilde{q} < 1; \\ \log\left(\frac{\tilde{r}-\tilde{a}}{1-\tilde{a}}\right), & \textit{if } \tilde{q} = 1. \end{cases} \tag{166}$$

Proof. We define $\varepsilon(p_i)$ as follows.

$$\varepsilon(p_i) = \begin{cases} \log\left(\frac{2p_i(1-\tilde{q})}{\sqrt{(1-p_i)^2+4p_i(1-\tilde{q})\left(\frac{1}{r^*(p_i,\tilde{q})}-p_i\tilde{q}\right)}-(1-p_i)}\right), & \text{if } 0 < \tilde{q} < 1; \\ \log\left(\frac{1-p_i}{\frac{1}{r^*(p_i,\tilde{q})}-p_i}\right), & \text{if } \tilde{q} = 1. \end{cases} \tag{167}$$

The solution to the minimization problem in (13) corresponds to the minimum of $\varepsilon(p_i)$ over $(0, 1]$.

We begin with the case where $0 < \tilde{q} \leq \tilde{a}/\tilde{r}$. In this setting, we have that $\tilde{r} \leq \tilde{a}/\tilde{q} \leq \tilde{a}/(p_i\tilde{q})$ so $r^*(p_i, \tilde{q}) = \tilde{a}/(p_i\tilde{q})$ for all p_i. By Lemma 2, $\varepsilon(p_i)$ is a decreasing function of p_i, so the minimum occurs at $p_i = 1$ and is, by Corollary 2,

$$\varepsilon(1) = \frac{1}{2}\log\left(\frac{1-\tilde{q}}{\frac{\tilde{q}}{\tilde{a}}-\tilde{q}}\right) = \frac{1}{2}\log\left(\frac{\tilde{a}(1-\tilde{q})}{\tilde{q}(1-\tilde{a})}\right). \tag{168}$$

We now consider the case where $\tilde{a}/\tilde{r} < \tilde{q} \leq 1$. Note that $\tilde{r} \geq \tilde{a}/(p_i\tilde{q})$ whenever $p_i \geq \tilde{a}/(\tilde{q}\tilde{r})$. Since in this setting $\tilde{a}/(\tilde{q}\tilde{r}) < 1$, the risk bound can be written in piece-wise form

$$r^*(\tilde{p}, q_i) = \begin{cases} \frac{\tilde{a}}{p_i\tilde{q}}, & \text{if } 0 < p_i < \frac{\tilde{a}}{\tilde{q}\tilde{r}}; \\ \tilde{r}, & \text{if } \frac{\tilde{a}}{\tilde{q}\tilde{r}} \leq p_i < 1. \end{cases} \tag{169}$$

Note that if $p_i \geq 1/(\tilde{r}\tilde{q})$, then $\tilde{r} \geq 1/(p_i\tilde{q})$. Thus, $r^*(p_i, \tilde{q}) \geq \tilde{r} \geq 1/(p_i\tilde{q})$ and so

$$\begin{aligned} &\sqrt{(1-p_i)^2 + 4p_i(1-\tilde{q})\left(\frac{1}{r^*(p_i,\tilde{q})} - p_i\tilde{q}\right)} - (1-p_i) \\ &\qquad \leq \sqrt{(1-p_i)^2 + 4p_i(1-\tilde{q})\left(p_i\tilde{q} - p_i\tilde{q}\right)} - (1-p_i) \qquad (170) \\ &\qquad = \sqrt{(1-p_i)^2} - (1-p_i) = 0. \qquad (171) \end{aligned}$$

Thus, $\varepsilon(p_i) = \infty$ for any $p_i \geq 1/(\tilde{r}\tilde{q})$. We may thus restrict our attention to $p_i < \min\{1, 1/(\tilde{r}\tilde{q})\}$.

Since $r^*(p_i, \tilde{q})$ is a continuous function of p_i and $\varepsilon(p_i)$ is a continuous function of $r^*(p_i, \tilde{q})$, it follows that $\varepsilon(p_i)$ is a continuous function of p_i. By Lemma 2, $\varepsilon(p_i)$ is a decreasing function of p_i for $0 < p_i < \tilde{a}/(\tilde{q}\tilde{r})$.

When $0 < \tilde{q} < 1/(\tilde{r}+1)$, note that $1/(\tilde{q}\tilde{r}) > (\tilde{r}+1)/\tilde{r} > 1$ and so $\min\{1, 1/(\tilde{r}\tilde{q})\} = 1$. By Lemma 3, $\varepsilon(p_i)$ is a decreasing function of p_i for $\tilde{a}/(\tilde{q}\tilde{r}) < p_i < 1$. Thus, the minimum must occur at $p_i = 1$ and is, by Corollary 2,

$$\varepsilon(1) = \frac{1}{2}\log\left(\frac{1-\tilde{q}}{\frac{1}{\tilde{r}}-\tilde{q}}\right). \tag{172}$$

When $\tilde{q} = 1/(\tilde{r}+1)$, again $1/(\tilde{q}\tilde{r}) = (\tilde{r}+1)/\tilde{r} > 1$. By Lemma 3, $\varepsilon(p_i)$ is flat for $\tilde{a}/(\tilde{q}\tilde{r}) < p_i < \min\{1, 1/(\tilde{r}\tilde{q})\}$ = 1. Thus, the minimum is shared by all points in this range and is equal to (172).

When $1/(\tilde{r}+1) < \tilde{q} < 1$, by Lemma 3, $\varepsilon(p_i)$ is an increasing function of p_i for $\tilde{a}/(\tilde{q}\tilde{r}) < p_i < \min\{1, 1/(\tilde{r}\tilde{q})\}$. Thus, the minimum must occur at $p_i = \tilde{a}/(\tilde{q}\tilde{r})$ and is

$$\varepsilon\left(\frac{\tilde{a}}{\tilde{q}\tilde{r}}\right) = \log\left(\frac{2\frac{\tilde{a}}{\tilde{q}\tilde{r}}(1-\tilde{q})}{\sqrt{(1-\frac{\tilde{a}}{\tilde{q}\tilde{r}})^2 + 4\frac{\tilde{a}}{\tilde{q}\tilde{r}}(1-\tilde{q})\left(\frac{1}{\tilde{r}} - \frac{\tilde{a}}{\tilde{q}\tilde{r}}\tilde{q}\right)} - (1-\frac{\tilde{a}}{\tilde{q}\tilde{r}})}\right) \tag{173}$$

$$= \log\left(\frac{2\tilde{a}(1-\tilde{q})}{\sqrt{(\tilde{q}\tilde{r}-\tilde{a})^2 + 4\tilde{a}\tilde{q}(1-\tilde{q})\,(1-\tilde{a})} - (\tilde{q}\tilde{r}-\tilde{a})}\right). \tag{174}$$

Finally, when $\tilde{q} = 1$ for $\tilde{a}/(\tilde{q}\tilde{r}) < p_i < \min\{1, 1/(\tilde{r}\tilde{q})\}$,

$$\varepsilon(p_i) = \log\left(\frac{1-p_i}{\frac{1}{\tilde{r}} - p_i}\right). \tag{175}$$

It was shown in (147) that $\varepsilon(p_i)$ is an increasing function of p_i for $\tilde{a}/\tilde{r} < p_i < \min\{1, 1/\tilde{r}\}$. The minimum then must occur at $p_i = \tilde{a}/\tilde{r}$ and is

$$\varepsilon\left(\frac{\tilde{a}}{\tilde{r}}\right) = \log\left(\frac{1-\frac{\tilde{a}}{\tilde{r}}}{\frac{1}{\tilde{r}} - \frac{\tilde{a}}{\tilde{r}}}\right) = \log\left(\frac{\tilde{r}-\tilde{a}}{1-\tilde{a}}\right). \tag{176}$$

□

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Our contribution is a novel, general framework for selecting ε in DP, which is stated in the abstract and introduction. We emphasize throughout that the framework requires some assumptions, which are provided at the beginning of Section 3.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The main limitation of our framework is the required assumptions, especially Assumption 2. We make an effort to call out that these assumptions are required for our framework to be used, while also noting that they are milder than assumptions in the existing literature on selecting ε.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: All theorems and lemmas are proved in Appendix E and all assumptions are clearly stated in Section 3.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: All equations required to reproduce the tables and figures are provided in the text. To find the ε implied by each profile, the minimization problem is stated in (13), closed forms are provided in Appendix A.2 where applicable, and code to solve the problem numerically is provided.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Code to produce all figures and tables in the text is provided.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: We present illustrations of the approach with results derived from mathematical properties of the framework rather than data-based experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: Our numerical results derive from calculations based on probabilistic properties of the framework, so error bars or other measures of statistical significance are unnecessary.

   Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: While we do not have any experiments, the code to produce all figures and tables can be run on a personal computer in a few seconds.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviewed and ensured that our paper meets the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

Justification: We demonstrate how our framework can help agencies better manage risk and utility trade-offs in private data releases, which can help agencies release information that better informs policy decisions. However, we do not discuss societal impacts beyond that notion.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks as it uses no confidential data.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.

- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.