Differential Privacy with Bounded Priors: Reconciling Utility and Privacy in Genome-Wide Association Studies

Florian Tramèr Zhicong Huang
Jean-Pierre Hubaux
School of IC, EPFL
firstname.lastname@epfl.ch

Erman Ayday*
Computer Engineering Department
Bilkent University
erman@cs.bilkent.edu.tr

ABSTRACT

Differential privacy (DP) has become widely accepted as a rigorous definition of data privacy, with stronger privacy guarantees than traditional statistical methods. However, recent studies have shown that for reasonable privacy budgets, differential privacy significantly affects the expected utility. Many alternative privacy notions which aim at relaxing DP have since been proposed, with the hope of providing a better tradeoff between privacy and utility.

At CCS'13, Li et al. introduced the membership privacy framework, wherein they aim at protecting against set membership disclosure by adversaries whose prior knowledge is captured by a family of probability distributions. In the context of this framework, we investigate a relaxation of DP, by considering prior distributions that capture more reasonable amounts of background knowledge. We show that for different privacy budgets, DP can be used to achieve membership privacy for various adversarial settings, thus leading to an interesting tradeoff between privacy guarantees and utility.

We re-evaluate methods for releasing differentially private χ^2-statistics in genome-wide association studies and show that we can achieve a higher utility than in previous works, while still guaranteeing membership privacy in a relevant adversarial setting.

Categories and Subject Descriptors

K.4.1 [**Computer and Society**]: Public Policy Issues—*Privacy*; C.2.0 [**Computer-Communication Networks**]: General—*Security and protection*; J.3 [**Life and Medical Sciences**]: Biology and genetics

Keywords

Differential Privacy; Membership Privacy; GWAS; Genomic Privacy; Data-Driven Medicine

1. INTRODUCTION

The notion of differential privacy, introduced by Dwork et al. [4, 5], provides a strong and rigorous definition of data privacy. A probabilistic mechanism $\mathcal{A}$ is said to satisfy ϵ-differential privacy (ϵ-DP), if for any two neighboring datasets T and T', the probability distribution of the outputs $\mathcal{A}(T)$ and $\mathcal{A}(T')$ differ at most by a multiplicative factor e^{ϵ}. Depending on the definition of neighboring datasets, we refer to either *unbounded*-DP or *bounded*-DP. Informally, satisfying differential privacy ensures that an adversary cannot tell with high confidence whether an entity t is part of a dataset or not, even if the adversary has complete knowledge over t's data, as well as over all the other entities in the dataset. The relevance of such a strong adversarial setting has been put into question because it seems unlikely, in a practical data setting, for an adversary to have such a high certainty about all entities. Alternative privacy definitions such as *differential-privacy under sampling* [13], *crowd-blending privacy* [8], *coupled-worlds privacy* [2], *outlier privacy* [15], *ϵ-privacy* [16], or *differential identifiability* [12] relax the adversarial setting of DP, with the goal of achieving higher utility.

This line of work is partially in response to the flow of recent results, for example in medical research, which show that satisfying differential privacy for reasonable privacy budgets leads to an significant drop in utility. For instance, Fredrikson et al. [6] investigate personalized warfarin dosing and demonstrate that for privacy budgets effective against a certain type of inference attacks, satisfying DP exposes patients to highly increased mortality risks. Similarly, studies on privacy in genome-wide association studies (GWAS) [10, 19, 22] consider differential privacy as a protective measure against an inference attack discovered by Homer et al. [9, 20]. These works show that for reasonably small values of ϵ, the medical utility is essentially null under DP, unless there is an access to impractically large patient datasets.

Membership Privacy.

We present an alternative characterization of differential privacy, by considering weaker adversarial models in the context of the positive membership-privacy (PMP) framework introduced by Li et al. [14]. Their privacy notion aims at preventing positive membership disclosure, meaning that an adversary should not be able to significantly increase his belief that an entity belongs to a dataset. The privacy guarantee is with respect to a distribution family $\mathbb{D}$, that captures an adversary's prior knowledge about the dataset. If

*Part of this work was done while the author was at EPFL.

CCS'15, October 12–16, 2015, Denver, Colorado, USA.
© 2015 ACM. ISBN 978-1-4503-3832-5/15/10 ...$15.00.
DOI: http://dx.doi.org/10.1145/2810103.2813610.

a mechanism $\mathcal{A}$ satisfies γ-positive membership-privacy under a family of distributions $\mathbb{D}$, denoted $(\gamma, \mathbb{D})$-PMP, then any adversary with a prior in $\mathbb{D}$ has a posterior belief upper-bounded in terms of the prior and the privacy parameter γ. The power of this framework lies in the ability to model different privacy notions, by considering different families of distributions capturing the adversary's prior knowledge. For instance, Li et al. show that ϵ-DP is equivalent to e^{ϵ}-PMP under a family of 'mutually independent distributions' (denoted either $\mathbb{D}_I$ for *unbounded*-DP or $\mathbb{D}_B$ for *bounded*-DP). Similarly, privacy notions such as differential identifiability or differential-privacy under sampling can also be seen as instantiations of the PMP framework for particular distribution families.

Bounded Adversarial Priors.

Our approach at relaxing the adversarial setting of DP is based on the observation that the families of mutually independent distributions $\mathbb{D}_I$ and $\mathbb{D}_B$ contain priors that assign arbitrarily high or low probabilities to all entities. This captures the fact that DP protects the privacy of an entity, even against adversaries with complete certainty about all other entities in the dataset, as well as some arbitrary (but not complete) certainty about the entity itself.

A natural relaxation we consider is to limit our adversarial model to mutually independent distributions that assign *bounded* prior probabilities to each entity. More formally, for constants $0 < a \leq b < 1$, we concentrate on adversaries with priors $p_t \in [a, b] \cup \{0, 1\}$ about the presence of each entity t in the dataset. In this setting, there are some entities (called *known* entities) for which the adversary knows apriori with absolute certainty whether they are in the dataset or not. For the remaining entities however (called *uncertain entities*), the adversary has some level of uncertainty about the entity's presence or absence from the dataset. In a sense, we consider what privacy guarantees a mechanism can provide for an uncertain entity, if the adversary has some limited amount of background knowledge about that entity. In contrast, DP asks for something much stronger, as it provides the same privacy guarantees for an entity, even if the adversary already has an arbitrarily high certainty about the entity's presence in the dataset.

Our main result shows that for a fixed privacy parameter ϵ, satisfying e^{ϵ}-PMP for adversaries with bounded priors requires less data perturbation than for the general families $\mathbb{D}_B$ and $\mathbb{D}_I$. More precisely, we prove that although ϵ-DP is necessary to guarantee e^{ϵ}-PMP for $\mathbb{D}_I$ and $\mathbb{D}_B$ (see [14]), a weaker level of ϵ'-DP (where $\epsilon' > \epsilon$) suffices to satisfy e^{ϵ}-PMP if the priors are bounded. Therefore, we introduce an alternative privacy-utility tradeoff, in which the data perturbation, and the utility loss, depend on the range of priors for which we guarantee a given level of PMP. This leads to an interesting model for the selection of the DP privacy parameter, in which we first identify a relevant adversarial setting and corresponding level of PMP, and then select the value ϵ such that these specific privacy guarantees hold.

Let's consider an interesting sub-case of our model of bounded prior distributions, where we let a get close to b; this corresponds to a setting where an adversary's prior belief about an entity's presence in the dataset tends to uniform, for those entities whose privacy is not already breached apriori. Although this adversarial model seems simplistic, we argue that certain relevant privacy threats, such as the attack on genomic studies by Homer et al. [9, 20], can be seen as particular instantiations of it. We show that protecting against such adversaries is, quite intuitively, much easier than against adversaries with unbounded priors. In Figure 1, we illustrate how the DP budget ϵ evolves, if our goal is to satisfy 2-PMP for priors ranging from a uniform belief of $\frac{1}{2}$ for each uncertain entity, to a general unbounded prior ($\mathbb{D}_B$ or $\mathbb{D}_I$). The figure should be read as follows: If the priors are arbitrary ($p_t \in [0, 1]$), then 2-PMP is guaranteed by satisfying $(\ln 2)$-DP. If the priors are uniformly $\frac{1}{2}$, then satisfying $(\ln 3)$-DP suffices. Note that for a prior of $\frac{1}{2}$, the definition of 2-PMP (see Definition 5) guarantees that the adversary's posterior belief that an uncertain entity is in the dataset is at most $\frac{3}{4}$.

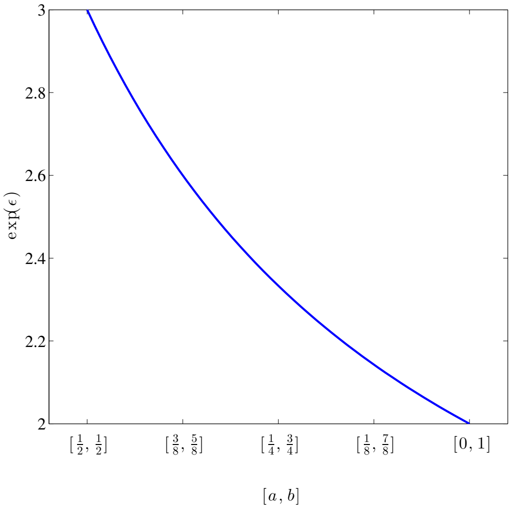


Figure 1: Level of ϵ-DP guaranteeing 2-PMP for the family of mutually independent distributions with priors bounded between a and b.

Result Assessment and Implications.

To assess the potential gain in utility of our relaxation, we focus on a particular application of DP, by re-evaluating the privacy protecting mechanisms in genome-wide association studies [10, 19, 22] for the release of SNPs with high χ^2-statistics. Our results show that, for a bounded adversarial model, we require up to 2500 fewer patients in the study, in order to reach an acceptable tradeoff between privacy and medical utility. As patient data is usually expensive and hard to obtain, this shows that a more careful analysis of the adversarial setting in a GWAS can significantly increase the practicality of known privacy preserving mechanisms.

As our theoretical results are not limited to the case of genomic studies, we believe that our characterization of DP for bounded adversarial models could be applied to many other scenarios, where bounded- or unbounded-DP has been considered as a privacy notion.

2. NOTATIONS AND PRELIMINARIES

We will retain most of the notation introduced for the membership-privacy framework in [14]. The universe of entities is denoted $\mathcal{U}$. An entity $t \in \mathcal{U}$ corresponds to a physical entity for which we want to provide some privacy-protection

List of symbols

$\mathcal{A}$	A privacy preserving mechanism
$\mathcal{U}$	The universe of entities
t	An entity in the universe $\mathcal{U}$
T	A subset of entities in $\mathcal{U}$ that make up the dataset
$\mathcal{D}$	A probability distribution over $2^{\mathcal{U}}$, representing the prior belief of some adversary about T
$\mathbf{T}$	A random variable drawn from $\mathcal{D}$ (the adversary's prior belief about T)
$\mathbb{D}$	A set of probability distributions
$\mathbb{D}_I$	The set of mutually independent distributions
$\mathbb{D}_B$	The set of bounded mutually independent distributions
$\mathbb{D}^{[a,b]}$	A subset of $\mathbb{D}$, in which all distributions assign priors in $[a,b] \cup \{0,1\}$ to all entities
$\mathbb{D}^a$	Equivalent to $\mathbb{D}^{[a,a]}$
ϵ	Privacy parameter for DP
γ	Privacy parameter for PMP

guarantees. A dataset is generated from the data associated with a subset of entities $T \subseteq \mathcal{U}$. By abuse of notation, we will usually simply denote the dataset as T. In order to model an adversary's prior belief about the contents of the dataset, we consider probability distributions $\mathcal{D}$ over $2^{\mathcal{U}}$ (the powerset of $\mathcal{U}$). From the point of view of the adversary, the dataset is a random variable $\mathbf{T}$ drawn from $\mathcal{D}$. Its prior belief that some entity t is in the dataset is then given by $\Pr_{\mathcal{D}}[t \in \mathbf{T}]$. In order to capture a range of adversarial prior beliefs, we consider a family of probability distributions. We denote a set of probability distributions by $\mathbb{D}$. Each distribution $\mathcal{D} \in \mathbb{D}$ corresponds to a particular adversarial prior we protect against. We denote a probabilistic privacy-preserving mechanism as $\mathcal{A}$. On a particular dataset T, the mechanism's output $\mathcal{A}(T)$ is thus a random variable. We denote by $\mathsf{range}(\mathcal{A})$ the set of possible values taken by $\mathcal{A}(T)$, for any $T \subseteq \mathcal{U}$.

2.1 Differential Privacy

Differential privacy provides privacy guarantees that depend solely on the privacy mechanism considered, and not on the particular dataset to be protected. Informally, DP guarantees that an entity's decision to add its data to a dataset (or to remove it) does not significantly alter the output distribution of the privacy mechanism.

Definition 1 (Differential Privacy [4, 5])**.** *A mechanism $\mathcal{A}$ provides ϵ-differential privacy if and only if for any two datasets T_1 and T_2 differing in a single element, and any $S \subseteq \mathsf{range}(\mathcal{A})$, we have*

$$\Pr\left[\mathcal{A}(T_1) \in S\right] \leq e^{\epsilon} \cdot \Pr\left[\mathcal{A}(T_2) \in S\right] \ . \tag{1}$$

Note that the above definition relies on the notion of datasets *differing in a single element*, also known as *neighboring* datasets. There exist two main definitions of neighboring datasets, corresponding to the notions of unbounded and bounded differential-privacy.

Definition 2 (Bounded DP [4])**.** *In bounded differential-privacy, datasets T_1 and T_2 are neighbors if and only if $|T_1| = |T_2| = k$ and $|T_1 \cap T_2| = k-1$. Informally, T_1 is obtained from T_2 by replacing one data entry by another.*

Definition 3 (Unbounded Differential-Privacy [5])**.** *In unbounded differential-privacy, datasets T_1 and T_2 are neighbors if and only if $T_1 = T_2 \cup \{t\}$ or $T_1 = T_2 \setminus \{t\}$, for some entity t. Informally, T_1 is obtained by either adding to, or removing an data entry from T_2.*

In this work, we consider two standard methods to achieve ϵ-DP, the so-called Laplace and exponential mechanisms. We first introduce the *sensitivity* of a function $f : 2^{\mathcal{U}} \rightarrow \mathbb{R}^n$; it characterizes the largest possible change in the value of f, when one data element is replaced.

Definition 4 (l_1-sensitivity [5])**.** *The l_1-sensitivity of a function $f : 2^{\mathcal{U}} \rightarrow \mathbb{R}^n$ is $\Delta f = \max_{T_1,T_2} ||f(T_1) - f(T_2)||_1$, where T_1 and T_2 are neighboring datasets.*

Laplace Mechanism.

If the mechanism $\mathcal{A}$ produces outputs in $\mathbb{R}^n$, the most straightforward method to satisfy DP consists in perturbing the output with noise drawn from the Laplace distribution. Let $\mathcal{A}$ be a mechanism computing a function $f : 2^{\mathcal{U}} \rightarrow \mathbb{R}^n$. Then, if on dataset T, $\mathcal{A}$ outputs $f(T) + \mu$, where μ is drawn from a Laplace distribution with mean 0 and scale $\frac{\Delta f}{\epsilon}$, then $\mathcal{A}$ satisfies ϵ-differential privacy [5].

Exponential Mechanism.

If $\mathcal{A}$ does not produce a numerical output, the addition of noise usually does not make sense. A more general mechanism guaranteeing ϵ-DP consists in defining a score function $q : T \times \mathsf{range}(\mathcal{A}) \rightarrow \mathbb{R}$ that assigns a value to each input-output pair of $\mathcal{A}$. On a dataset T, the exponential mechanism samples an output $r \in \mathsf{range}(\mathcal{A})$ with probability proportional to $\exp(\frac{q(T,r)\epsilon}{2\Delta q})$, which guarantees ϵ-DP [17].

2.2 Positive Membership-Privacy

In this subsection, we give a review of the membership-privacy framework from [14] and its relation to differential-privacy. Readers familiar with this work can skip directly to Section 3, where we introduce and discuss our relaxed adversarial setting.

The original membership-privacy framework is comprised of both positive and negative membership-privacy. In this work, we are solely concerned with positive membership-privacy (PMP). This notion protects against a type of re-identification attack called *positive membership disclosure*, where the output of the mechanism $\mathcal{A}$ significantly increases an adversary's belief that some entity belongs to the dataset. Adversaries are characterized by their prior belief over the contents of the dataset T. A mechanism $\mathcal{A}$ is said to satisfy positive membership-privacy for a given prior distribution, if after the adversary sees the output of $\mathcal{A}$, its posterior belief about an entity belonging to a dataset is not significantly larger than its prior belief.

Note that although differential privacy provides seemingly strong privacy guarantees, it does not provide PMP for adversaries with arbitrary prior beliefs. It is well known that data privacy against arbitrary priors cannot be guaranteed if some reasonable level of utility is to be achieved. This fact, known as the *no-free-lunch-theorem*, was first introduced by Kifer and Machanavajjhala [11], and reformulated by Li et al. [14] as part of their framework. We now give the formal definition of γ-positive membership-privacy under a family of prior distributions $\mathbb{D}$, which we denote as $(\gamma, \mathbb{D})$-PMP.

Definition 5 (Positive Membership-Privacy [14])**.** *A mechanism $\mathcal{A}$ satisfies γ-PMP under a distribution family $\mathbb{D}$, where $\gamma \geq 1$, if and only if for any $S \subseteq \mathsf{range}(\mathcal{A})$, any distribution $\mathcal{D} \in \mathbb{D}$, and any entity $t \in \mathcal{U}$, we have*

$$\Pr_{\mathcal{D}|\mathcal{A}}[t \in \mathbf{T} \mid \mathcal{A}(\mathbf{T}) \in S] \leq \gamma \Pr_{\mathcal{D}}[t \in \mathbf{T}] \tag{2}$$

$$\Pr_{\mathcal{D}|\mathcal{A}}[t \notin \mathbf{T} \mid \mathcal{A}(\mathbf{T}) \in S] \geq \frac{1}{\gamma} \Pr_{\mathcal{D}}[t \notin \mathbf{T}] \ . \tag{3}$$

By some abuse of notation, we denote by S the event $\mathcal{A}(\mathbf{T}) \in S$ and by t the event $t \in \mathbf{T}$. When $\mathcal{D}$ and $\mathcal{A}$ are obvious from context, we reformulate (2), (3) as

$$\Pr[t \mid S] \leq \gamma \Pr[t] \tag{4}$$

$$\Pr[\neg t \mid S] \geq \frac{1}{\gamma} \Pr[\neg t] \ . \tag{5}$$

Together, theses inequalities are equivalent to

$$\Pr[t \mid S] \leq \min\left(\gamma \Pr[t], \frac{\gamma - 1 + \Pr[t]}{\gamma}\right) \ . \tag{6}$$

The privacy parameter γ in PMP is somewhat analogous to the parameter e^ϵ in DP (we will see that the two privacy notions are equivalent for a particular family of prior distributions). Note that the smaller γ is, the closer the adversary's posterior belief is to its prior belief, implying a small knowledge gain. Thus, the strongest privacy guarantees correspond to values of γ close to 1.

Having defined positive membership-privacy, we now consider efficient methods to guarantee this notion of privacy, for various distribution families. A simple sufficient condition on the output of the mechanism $\mathcal{A}$, which implies PMP, is given by Li et al. in the following lemma.

Lemma 1 ([14])**.** *If for any distribution $\mathcal{D} \in \mathbb{D}$, any output $S \subseteq \mathsf{range}(\mathcal{A})$ and any entity t for which $0 < \Pr_{\mathcal{D}}[t] < 1$, the mechanism $\mathcal{A}$ satisfies*

$$\Pr[S \mid t] \leq \gamma \cdot \Pr[S \mid \neg t] \ ,$$

then $\mathcal{A}$ provides $(\gamma, \mathbb{D})$-PMP.

Notice the analogy to differential privacy here, in the sense that the above condition ensures that the probabilities of $\mathcal{A}$ producing an output, given the presence or absence of a particularly data entry, should be close to each other.

Relation to Differential Privacy.

One of the main results of [14] shows that differential privacy is equivalent to PMP under a particular distribution family. We will be primarily concerned with bounded DP, as it is the privacy notion generally used for the genome-wide association studies we consider in Section 4. Our main results also apply to unbounded DP and we discuss this relation in Section 5. Before presenting the main theorem linking the two privacy notions, we introduce the necessary distribution families.

Definition 6 (Mutually-Independent Distributions (MI) [14])**.** *The family $\mathbb{D}_I$ contains all distributions characterized by assigning a probability p_t to each entity t such that the probability of a dataset T is given by*

$$\Pr[T] = \prod_{t \in T} p_t \cdot \prod_{t \notin T} (1 - p_t) \ . \tag{7}$$

Definition 7 (Bounded MI Distributions (BMI) [14])**.** *A BMI distribution is the conditional distribution of a MI distribution, given that all datasets with non-zero probability have the same size. The family $\mathbb{D}_B$ contains all such distributions.*

The following result, used in the proof of Theorem 4.8 in [14] will be useful when we consider relaxations of the family $\mathbb{D}_B$ in Section 3.

Lemma 2 ([14])**.** *If $\mathcal{A}$ satisfies ϵ-bounded DP, then for any $\mathcal{D} \in \mathbb{D}_B$ we have*

$$\frac{\Pr[S \mid t]}{\Pr[S \mid \neg t]} \leq e^\epsilon \ .$$

Note that together with Lemma 1, this result shows that ϵ-bounded differential-privacy implies e^ϵ-positive membership-privacy under $\mathbb{D}_B$. Li et al. prove that the two notions are actually equivalent.

Theorem 1 ([14])**.** *A mechanism $\mathcal{A}$ satisfies ϵ-bounded DP if and only if it satisfies $(e^\epsilon, \mathbb{D}_B)$-PMP.*

This equivalence between ϵ-bounded DP and e^ϵ-PMP under $\mathbb{D}_B$ will be the starting point of our relaxation of differential privacy. Indeed, we will show that for certain subfamilies of $\mathbb{D}_B$, we can achieve e^ϵ-PMP even if we only provide a weaker level of differential privacy. In this sense, we will provide a full characterization of the relationship between the privacy budget of DP and the range of prior beliefs for which we can achieve e^ϵ-PMP.

3. PMP FOR BOUNDED PRIORS

The result of Theorem 1 provides us with a clear characterization of positive membership-privacy under the family $\mathbb{D}_B$. We now consider the problem of satisfying PMP for different distribution families. In particular, we are interested in protecting our dataset against adversaries weaker than those captured by $\mathbb{D}_B$, meaning adversaries with less background knowledge about the dataset's contents. Indeed, as the prior belief of adversaries considered by DP has been argued to be unreasonably strong for most practical settings, our goal is to consider a restricted adversary, with a more plausible level of background knowledge.

One reason to consider a weaker setting than DP's adversarial model, is that mechanisms that satisfy DP for small values of ϵ have been shown to provide rather disappointing utility in practice. Examples of studies, where DP offers a poor privacy-utility tradeoff, are numerous in medical applications such as genome-wide association studies [10, 19, 22] or personalized medicine [6]. Indeed, many recent results have shown that the amount of perturbation introduced by appropriate levels of DP on such datasets renders most statistical queries useless. We will show that when considering more reasonable adversarial settings, we can achieve strong membership-privacy guarantees with less data perturbation, thus leading to a possibly better privacy-utility tradeoff.

3.1 A Relaxed Threat Model

As illustrated by Theorem 1, differential privacy guarantees positive-membership privacy against adversaries with a prior in $\mathbb{D}_B$. Thus, in the context of protection against membership disclosure, the threat model of differential privacy considers adversaries with the following capabilities.

1. The adversary knows the size of the dataset N.
2. All entities are considered independent, conditioned on the dataset having size N.
3. There are some entities for which the adversary knows with absolute certainty whether they are in the dataset or not ($\Pr[t] \in \{0, 1\}$).
4. For all other entities, the adversary may have an arbitrary prior belief $0 < \Pr[t] < 1$ that the entity belongs to the dataset.

In our threat model, we relax capability 4). We first consider each capability separately and discuss why it is (or is not) a reasonable assumption for realistic adversaries.

Knowledge of N.

Bounded-DP inherently considers neighboring datasets of fixed size. It is preferably used in situations where the size of the dataset is public and fixed, an example being the genome-wide association studies we discuss in Section 4. In contrast, unbounded-DP is used in situations where the size of the dataset is itself private. Our results apply in both cases (see Section 5 for a discussion of unbounded-DP).

Independence of Entities.

As we have seen in Theorem 1 (and will see in Theorem 3 for unbounded-DP), a differentially-private mechanism guarantees that an adversary's posterior belief will be within a given multiplicative factor of its prior, exactly when the adversary's prior is a (bounded) mutually independent distribution. In this work, we focus on a relaxation of DP within the PMP framework, and thus model our adversary's prior belief as a subfamily of either $\mathbb{D}_B$ or $\mathbb{D}_I$.

Known Entities.

It is reasonable to assume that an adversary may know with certainty whether some entities belong to the dataset or not, because these entities either willingly or unwillingly disclosed their (non)-membership (the adversary itself may be an entity of the universe). Note that for such entities with prior 0 or 1, perfect PMP with $\gamma = 1$ is trivially satisfied, since the adversary's posterior does not differ from its prior. As the privacy of these entities is already breached *a priori*, the privacy guarantees of $\mathcal{A}$ should be considered only with respect to those entities whose privacy still can be protected. Because all entities are considered independent, we may assume that the adversary knows about some entities' presence in the dataset, but that some uncertainty remains about others.

Unknown Entities.

A distribution $\mathcal{D} \in \mathbb{D}_B$ can assign to each uncertain entity a prior probability arbitrarily close to 0 or 1. This means that when providing positive membership-privacy under $\mathbb{D}_B$, we are considering adversaries that might have an extremely high prior confidence about whether *each* user's data is contained in the dataset or not. In this sense, the family $\mathbb{D}_B$ corresponds to an extremely strong adversarial setting, as it allows for adversaries with arbitrarily high prior beliefs about the contents of a dataset.

Yet, while it is reasonable to assume that the adversary may know for certain whether some entities are part of the dataset or not, it seems unrealistic for an adversary to have high confidence about its belief for *all* entities, *a priori*. As we will see, guaranteeing membership privacy for those entities for which an adversary has high confidence *a priori* ($\Pr[t]$ close to 0 or 1), requires the most data perturbation. Thus, when protecting against adversaries with priors in $\mathbb{D}_B$, we are degrading our utility in favor of protection for entities whose membership privacy was already severely compromised to begin with. In our alternative threat model, we focus on protecting those entities whose presence in the dataset remains highly uncertain to the adversary prior to releasing the output of $\mathcal{A}$. As we will see in Section 3.4, our mechanisms still guarantee some weaker level of protection against the full set of adversaries with priors in $\mathbb{D}_B$.

3.2 Our Results

Our natural relaxation of DP's adversarial model consists in restricting ourselves to adversaries with a prior belief about uncertain entities bounded away from 0 and 1. Such an adversary thus may know for certain whether some entities are in the dataset or not, because they unwillingly or willingly disclosed this information to the adversary. For the remaining entities however, the adversary has some minimal level of uncertainty about the entity's presence or absence from the dataset, which appears to be a reasonable assumption to make in practice. We will consider the subfamily of $\mathbb{D}_B$, consisting of all BMI distributions for which the priors $\Pr[t]$ are either $0, 1$ or bounded away from 0 and 1. This distribution family is defined as follows.

Definition 8 (Restricted[1] BMI Distributions)**.** *For $0 < a \leq b < 1$, the family $\mathbb{D}_B^{[a,b]} \subset \mathbb{D}_B$ contains all BMI distributions for which $\Pr[t] \in [a, b] \cup \{0, 1\}$, for all entities t. If $a = b$, we simply denote the family as $\mathbb{D}_B^a$.*

Our goal is to show that in this weaker adversarial setting, we can guarantee PMP with parameter γ, while satisfying a weaker form of privacy than $(\ln \gamma)$-DP.

We first show that the adversaries with arbitrarily low or high priors are, rather intuitively, the hardest to protect against. More formally, we show that when guaranteeing $(\gamma, \mathbb{D}_B)$-PMP, inequalities (2) and (3) are only tight for priors approaching 0 or 1. For each entity t, we can compute a tight privacy parameter $\gamma(t) \leq \gamma$, whose value depends on the prior $\Pr[t]$. When considering an adversary with a prior belief in $\mathbb{D}_B^{[a,b]}$, we will see that $\gamma(t) < \gamma$ for all entities t, which shows that we can achieve tighter positive membership-privacy guarantees in our relaxed adversarial model. We formalize these results in the following lemma.

Lemma 3. *If a mechanism $\mathcal{A}$ satisfies $(\gamma, \mathbb{D}_B)$-PMP, then $\Pr[t \mid S] \leq \gamma(t) \cdot \Pr[t]$ and $\Pr[\neg t \mid S] \geq \frac{\Pr[\neg t]}{\gamma(t)}$, where*

$$\gamma(t) = \begin{cases} 1 & \textit{if } \Pr[t] \in \{0, 1\} \\ \max\{(\gamma-1)\Pr[t]+1, \ \frac{\gamma}{(\gamma-1)\Pr[t]+1}\} & \textit{otherwise.} \end{cases}$$

Proof. If $\Pr[t] \in \{0, 1\}$, $\gamma(t) = 1$ and the lemma trivially holds. If $0 < \Pr[t] < 1$, Bayes' theorem gives us

$$\Pr[t \mid S] = \frac{1}{1 + \frac{\Pr[S|\neg t]}{\Pr[S|t]} \frac{\Pr[\neg t]}{\Pr[t]}} \ . \tag{8}$$

[1] Although we talk about adversaries with bounded priors, we use the term *restricted* instead of *bounded* here, as $\mathbb{D}_B$ already denotes the family of bounded MI distributions in [14].

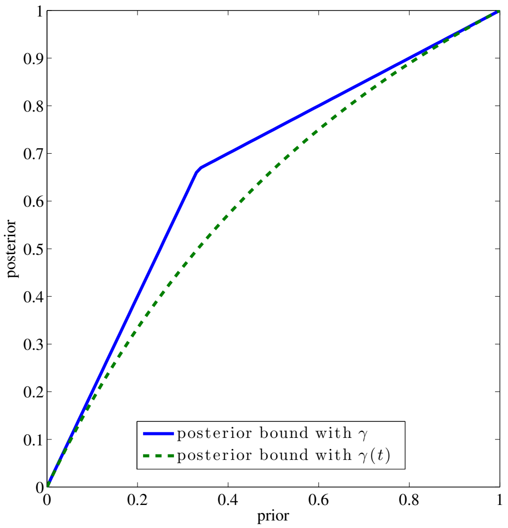


Figure 2: Bounds on an adversary's posterior belief when satisfying $(2, \mathbb{D}_B)$-PMP.

By Theorem 1, we know that providing $(\gamma, \mathbb{D}_B)$-PMP is equivalent to satisfying $(\ln \gamma)$-bounded DP. By Lemma 2, we then get

$$\Pr[t \mid S] \leq \frac{1}{1 + \gamma^{-1}\frac{1-\Pr[t]}{\Pr[t]}} = \frac{\gamma \cdot \Pr[t]}{(\gamma - 1)\Pr[t] + 1} \tag{9}$$

$$\Pr[\neg t \mid S] \geq \frac{1}{1 + \gamma\frac{\Pr[t]}{\Pr[\neg t]}} = \frac{\Pr[\neg t]}{(\gamma - 1)\Pr[t] + 1} \ . \tag{10}$$

□

From this lemma, we get that $\gamma(t) < \gamma$ for all entities for which $0 < \Pr[t] < 1$. Thus, as mentioned previously, $(\gamma, \mathbb{D}_B)$-PMP actually gives us a privacy guarantee stronger than the bounds (2) and (3), for all priors bounded away from 0 or 1. To illustrate this, Figure 2 plots the two different bounds on the posterior probability, when satisfying $(2, \mathbb{D}_B)$-PMP.

Let $\mathcal{A}$ be a mechanism satisfying $(\gamma, \mathbb{D}_B)$-PMP. If we were to consider only those distributions in $\mathbb{D}_B$ corresponding to prior beliefs bounded away from 0 and 1, then $\mathcal{A}$ would essentially satisfy PMP for some privacy parameter larger than γ. This privacy gain can be quantified as follows. From Lemma 3, we immediately see that if we satisfy $(\gamma, \mathbb{D}_B)$-PMP, then we also satisfy $(\gamma', \mathbb{D}_B^{[a,b]})$-PMP, where

$$\gamma' = \max_{t \in \mathcal{U}} \gamma(t) = \max\left((\gamma - 1)b + 1,\ \frac{\gamma}{(\gamma - 1)a + 1}\right) \ . \tag{11}$$

As $\gamma' < \gamma$, this result shows (quite unsurprisingly) that if we consider a weaker adversarial model, our privacy guarantee increases. Conversely, we now show that for a fixed privacy level, the relaxed adversarial model requires less data perturbation. Suppose we fix some positive membership-privacy parameter γ. We know that to provide $(\gamma, \mathbb{D}_B)$-PMP, we have to satisfy $(\ln \gamma)$-DP. However, our goal here is to satisfy $(\gamma, \mathbb{D}_B^{[a,b]})$-PMP for a tight value of γ. The following theorem shows that a sufficient condition to protect positive membership-privacy against a bounded adversary is to provide a weaker level of differential privacy.

Theorem 2. *A mechanism $\mathcal{A}$ satisfies $(\gamma, \mathbb{D}_B^{[a,b]})$-PMP, for some $0 < a \leq b < 1$, if $\mathcal{A}$ satisfies ϵ-bounded DP, where*

$$e^\epsilon = \begin{cases} \min\left(\frac{(1-a)\gamma}{1-a\gamma},\ \frac{\gamma+b-1}{b}\right) & \text{if } a\gamma < 1, \\ \frac{\gamma+b-1}{b} & \text{otherwise} \ . \end{cases}$$

Proof. Recall that satisfying ϵ-bounded differential privacy is equivalent to satisfying $(e^\epsilon, \mathbb{D}_B)$-PMP. Using (11), we want

$$\gamma = \max\left((e^\epsilon - 1)b + 1,\ \frac{e^\epsilon}{(e^\epsilon - 1)a + 1}\right) \ . \tag{12}$$

Solving for ϵ yields the desired result. □

Note that when $a\gamma \geq 1$, the first condition of PMP, namely $\Pr[t \mid S] \leq \gamma \Pr[t]$ is trivially satisfied. Thus, in this case we have to satisfy only the second condition, $\Pr[\neg t \mid S] \leq \frac{\Pr[\neg t]}{\gamma}$, which is satisfied if $\gamma = (e^\epsilon - 1)b + 1$. We thus arrive at a full characterization of the level of differential privacy to satisfy, if we wish to guarantee a certain level of positive membership-privacy for subfamilies of $\mathbb{D}_B$. For a fixed level of privacy γ and $0 < a \leq b < 1$, protecting against adversaries from a family $\mathbb{D}_B^{[a,b]}$ will correspond to a weaker level of differential privacy, and thus to less perturbation of the mechanism's outputs, compared to the distribution family $\mathbb{D}_B$. Therefore, by considering a more restricted adversarial setting, we could indeed reach a higher utility for a constant level of protection against positive membership disclosure.

These results lead to the following simple model for the selection of an appropriate level of differential privacy, in a restricted adversarial setting.

Selecting a level of DP

```
1: Identify a practically significant adversarial model captured by some distribution family $\mathbb{D}_B^{[a,b]}$.
2: Select a level $\gamma$ of PMP, providing appropriate bounds on the adversary's posterior belief.
3: Use Theorem 2 to get the corresponding level of DP.
```

As an example, assume a PMP parameter of 2 is considered to be a suitable privacy guarantee. If our adversarial model is captured by the family $\mathbb{D}_B$, then $(\ln 2)$-DP provides the necessary privacy. However, if a reasonable adversarial setting is the family $\mathbb{D}_B^{0.5}$, then the same privacy guarantees against membership disclosure are obtained by satisfying $(\ln 3)$-DP, with significantly less data perturbation.

3.3 Selecting the Bounds $[a, b]$ in Practice

Selecting appropriate bounds $[a, b]$ on an adversary's prior belief (about an individual's presence in a dataset) is primordial for our approach, yet might prove to be a difficult task in practice. One possibility is to focus on privacy guarantees in the presence of a particular identified adversarial threat. In Section 4.2, we will consider a famous attack on genome-wide association studies, and show how we can define bounds on the adversary's prior, in the presumed threat model. Such bounds are inherently heuristic, as they derive from a particular set of assumptions about the adversary's power, that might fail to be met in practice. However, we will show in Section 3.4, that our methods also guarantee some (smaller) level of privacy against adversaries whose prior beliefs fall outside of the selected range.

Finally, another use-case of our approach is for obtaining *upper bounds* on the utility that a mechanism may achieve, when guaranteeing γ-PMP against a so-called *uninformed* adversary. If the dataset size N and the size of the universe $\mathcal{U}$ are known, such an adversary *a priori* considers all individuals as part of the dataset with equal probability $\frac{N}{|\mathcal{U}|}$.

3.4 Risk-Utility Tradeoff

We have shown that focusing on a weaker adversary leads to higher utility, yet we must also consider the increased privacy risk introduced by this relaxation. Suppose our goal is to guarantee e^{ϵ}-PMP. If we consider the adversarial model captured by the full family $\mathbb{D}_B$, $\mathcal{A}$ must satisfy ϵ-DP. If we instead focus on the relaxed family $\mathbb{D}_B^{[a,b]}$, it suffices to guarantee ϵ'-DP, where ϵ' is obtained from Theorem 2.

Now suppose our mechanism satisfies $(e^{\epsilon}, \mathbb{D}_B^{[a,b]})$-PMP, but there actually is an entity for which the adversary has a prior $\Pr[t] \notin ([a,b] \cup \{0,1\})$. Although our mechanism will not guarantee that conditions (2) and (3) hold for this entity, a weaker protection against membership disclosure still holds. Indeed, since our mechanism satisfies ϵ'-DP, it also satisfies $(e^{\epsilon'}, \mathbb{D}_B)$-PMP by Theorem 1, and thus guarantees that bounds (2) and (3) will hold with a factor of $e^{\epsilon'}$, rather than e^{ϵ}. In conclusion, satisfying ϵ-DP corresponds to satisfying e^{ϵ}-PMP for all entities, regardless of the adversary's prior. Alternatively, satisfying ϵ'-DP is sufficient to guarantee e^{ϵ}-PMP for those entities for which the adversary has a bounded prior $\Pr[t] \in [a,b] \cup \{0,1\}$, and a weaker level of $e^{\epsilon'}$-PMP for entities whose membership privacy was already severely compromised to begin with.

3.5 Relation to Prior Work

A number of previous relaxations of differential privacy's adversarial model have been considered. We discuss the relations between some of these works and ours in this section.

A popular line of work considers *distributional* variants of differential privacy, where the dataset is assumed to be randomly sampled from some distribution known to the adversary. Works on *Differential-Privacy under Sampling* [13], *Crowd-Blending Privacy* [8], *Coupled-Worlds Privacy* [2] or *Outlier Privacy* [15] have shown that if sufficiently many users are *indistinguishable* by a mechanism, and this mechanism operates on a dataset obtained through a *robust* sampling procedure, differential privacy can be satisfied with only little data perturbation. Our work differs in that we make no assumptions on the indistinguishability of different entities, and that our aim is to guarantee membership privacy rather than differential privacy. Another main difference is in the prior distributions of the adversaries that we consider. Previous works mainly focus on the unbounded-DP case, and thus are not directly applicable to situations where the size of the dataset is public. Furthermore, previously considered adversarial priors are either uniform [13, 2] or only allow for a fixed number of known entities [8, 15]. Finally, very few results are known on how to design general mechanisms satisfying distributional variants of DP. In our work, we show how different levels of DP, for which efficient mechanisms are known, can be used to guarantee PMP for various adversarial models. Alternatively, *Differential Identifiability* [12] was shown in [14] to be equivalent to PMP under a family of prior distributions slightly weaker than the ones we introduce here, namely where all entities have a prior $\Pr[t] \in \{0, \beta\}$ for some fixed β.

4. EVALUATION

Having provided a theoretical model for the characterization of DP for adversaries with bounded priors, we now evaluate the new tradeoff between privacy and utility that we introduce when considering adversarial models captured by a family $\mathbb{D}_B^{[a,b]}$. We can view an adversary with a prior in this family as having complete certainty about the size of the dataset, as well as some degree of uncertainty about its contents. Scenarios that nicely fit this model, and have been gaining a lot of privacy-focused attention recently, are genome-wide association studies (GWAS). We will use this setting as a case study for the model we propose for the selection of an appropriate DP parameter.

4.1 Genome-Wide Association Studies

Let us begin with some genetic background. The human genome consists of about 3.2 billion **base pairs**, where each base pair is composed of two **nucleobases** (**A**,**C**,**G** or **T**). Approximately 99.5% of our genome is common to all human beings. In the remaining part of our DNA, a **single nucleotide polymorphism** (**SNP**) denotes a type of genetic variation occurring commonly in a population. A SNP typically consists of a certain number of possible nucleobases, also called **alleles**. An important goal of genetic research is to understand how these variations in our **genotypes** (our genetic material), affect our **phenotypes** (any observable trait or characteristic, a particular disease for instance).

We are concerned with SNPs that consist of only two alleles and occur on a particular chromosome. Each such SNP thus consists of two nucleobases, one on each chromosome. An example of a SNP is given in Figure 3. In a given population, the **minor allele frequency** (**MAF**) denotes the frequency at which the least common of the two alleles occurs on a particular SNP.

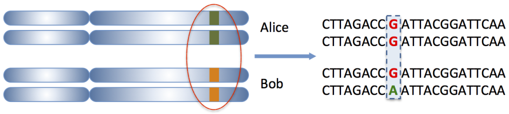


Figure 3: Example of a SNP. Alice has two G alleles on this fragment and Bob has one G allele and one A allele.

We use the standard convention to encode the value of a SNP as the number of minor alleles it contains. As an example, if a SNP has alleles A and G, and A is the minor allele, then we encode SNP GG as 0, SNPs AG and GA as 1, and SNP AA as 2. The MAF corresponds to the frequency at which SNP values 1 or 2 occur.

Genome-wide association studies (**GWAS**) are a particular type of **case-control** studies. Participants are divided into two groups, a **case group** containing patients with a particular phenotype (a disease for instance) and a **control group**, containing participants without the attribute. For each patient, we record the values of some particular SNPs, in order to determine if any DNA variation is associated with the presence of the studied phenotype. If the value of a SNP appears to be correlated (negatively or positively) to the phenotype, we say that the SNP is **causative**, or **associated** with the phenotype.

A standard way to represent this information is through a **contingency table** for each of the considered SNPs. For a particular SNP, this table records the number of cases and controls having a particular SNP value. An example of such a table is given hereafter, for a GWAS involving 100 cases and 100 controls. From this table, we can, for instance, read that 70% of the cases have no copy of the minor allele. We can also compute the MAF of the SNP as $\frac{40+2\cdot 50}{2\cdot 200} = 0.35$.

SNP value	Cases	Controls	Total
0	70	40	110
1	10	30	40
2	20	30	50
Total	100	100	200

Table 1: Contingency table of one SNP, for a GWAS with 100 cases and 100 controls.

The interested reader may find additional information on genomics as well as on genome privacy and security research at the community website[2] maintained by our group.

4.2 Homer's Attack and Adversarial Model

The ongoing research on applying differential privacy to GWAS has been primarily motivated by an attack proposed by Homer et al. [9]. In this attack, the adversary is assumed to have some knowledge about an entity's genetic profile, and wants to determine if this entity belongs to the case group or not. Towards this end, the adversary measures the distance between the entity's SNP values and the allele frequencies reported for the case group, or some reference population. It has been shown that other aggregate statistics, such as p-values or SNP correlation scores, could be used to construct similar or even stronger attacks [20].

Unsurprisingly, privacy mechanisms based on DP have been proposed to counter these attacks [10, 19, 22], as they guarantee that an entity's presence in the dataset will not significantly affect the output statistics. However, the adversarial model considered here is quite different from the one DP protects against. Indeed, all these attacks assume some prior knowledge about an entity's genomic profile, but not about the entity's presence or absence from the case group. Actually, the adversary makes no assumptions on the presence or absence of any entity from the case group, and it is absolutely not assumed to have complete knowledge about the data of all but one of the entities. This attack thus appropriately fits into our relaxed adversarial setting, where we consider an adversary with bounded prior knowledge. From the results of Section 3, we know that protecting membership disclosure against such adversaries can be achieved with much weaker levels of DP.

In the following, we consider a genome-wide association study with N patients. It is generally recommended ([18]) that the number of cases and controls be similar. We thus focus on studies with $\frac{N}{2}$ cases and $\frac{N}{2}$ controls. The cases suffer from a particular genetic disease, and the goal of the study is to find associated SNPs by releasing some aggregate statistics over all participants. We assume that the adversary knows the value N (which is usually reported by the study). In the adversarial model considered by DP, we would assume the adversary to have complete knowledge about all but one of the entities in the case group. We will consider a weaker setting here, which includes the adversarial model of Homer's attack [9]. The adversary is assumed to know the identity of the study participants, and possibly the disease status of some of them, but has no additional information on whether other entities were part of the case or control group. In regard to the attacks discussed previously, we will limit the adversary's capability of asserting the membership of an entity to the case group, and thus his disease status.

Suppose the adversary already breached the privacy of a small number m_1 of the cases and m_2 of the controls. In this case, the adversary's prior belief about some other entity's presence in the case group is $\frac{N/2-m_1}{N-m_1-m_2}$. In the following, we assume that $m_1 \approx m_2$ and thus that the adversary's prior can be modeled by the family $\mathbb{D}_B^{0.5}$. As we discussed in Section 3.4, our mechanisms will still provide some smaller level of security against adversaries with more general priors.

More generally, if we have N_1 cases and N_2 controls, we can consider a similar adversarial model with a prior belief of $\frac{N_1}{N_1+N_2}$ that an entity belongs to the case group.

4.3 A Simple Counting Query

We first consider a simple *counting query*. While the following example has little practical significance in a GWAS, it is an interesting and simple toy-example illustrating the usability and power of the model we derived in Section 3.

Let $\mathcal{A}$ and $\mathcal{A}'$ be mechanisms computing the number of patients in the case group whose SNP value is 0. Under bounded-DP, the sensitivity of this query is 1. Suppose we want to guarantee $(\gamma, \mathbb{D}_B)$-PMP for $\mathcal{A}$, and $(\gamma, \mathbb{D}_B^{0.5})$-PMP for $\mathcal{A}'$. In the first case, this is equivalent to satisfying ϵ-DP, for $\epsilon = \ln(\gamma)$. In the bounded adversarial model, we have shown in Theorem 2 that it is sufficient to satisfy ϵ'-DP, for an $\epsilon' > \ln(\gamma)$.

To satisfy DP, and therefore PMP, we add Laplacian noise to the true count value. We define the utility of our mechanism as the precision of the count, after application of the privacy mechanism. More formally, if the true count is C and the noisy output count is $\hat{C}$, then we are interested in the expected error $\mathbb{E}[|\hat{C} - C|]$. When satisfying ϵ-DP, we have that $\hat{C} = C + \mu$, where μ is drawn from a Laplace distribution with mean 0 and scale ϵ^{-1}. Thus, we have that

$$\mathbb{E}[|\hat{C} - C|] = \mathbb{E}[|\mu|] = \epsilon^{-1} \ . \tag{13}$$

As a concrete example of the differences in utility between $\mathcal{A}$ and $\mathcal{A}'$, we vary the PMP parameter γ and plot the expected error of the count in Figure 4. As we can see, $\mathcal{A}'$ gives significantly more precise outputs than $\mathcal{A}$, when the two mechanisms provide the same positive membership-privacy guarantees in their respective adversarial settings. Note that for an adversary with prior $\mathbb{D}_B^{0.5}$, and PMP parameter $\lambda = 2$, seeing the output of $\mathcal{A}'$ yields a posterior belief of at most $\frac{3}{4}$, that a particular entity is in the case group. This simple example shows that by focusing on a bounded adversarial model, protecting against membership disclosure can be achieved while retaining significantly higher utility, compared to the original adversarial setting.

4.4 Releasing Causative SNPs

A typical GWAS aims at uncovering SNPs that are associated with some disease. A standard and simple method consists in computing the χ^2-statistics of the contingency table of each SNP. Assume that the genomes of people participating in the GWAS are uncorrelated (a necessary assumption for χ^2-statistics). For a SNP unrelated to the disease, we expect any SNP value to appear in the case group as often

[2] `https://genomeprivacy.org`

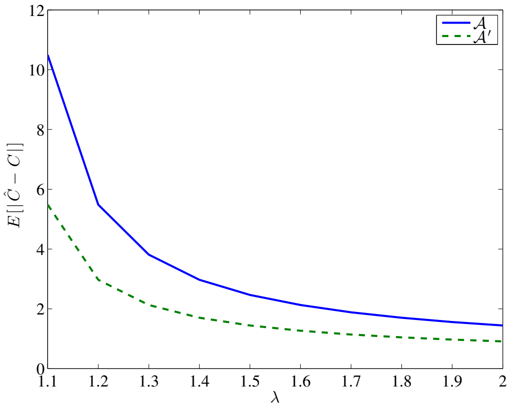


Figure 4: Expected error of the counting query, for privacy mechanisms $\mathcal{A}$ and $\mathcal{A}'$ satisfying $(\lambda, \mathbb{D}_B)$-PMP and $(\lambda, \mathbb{D}_B^{0.5})$-PMP respectively.

as in the control group. The χ^2-statistic measures how much the true values diverge from this expected **null hypothesis**. The higher the statistic is, the more likely it is that the SNP and disease status are correlated. Equivalently, we can compute the **p-values** that correspond to the χ^2-statistics.

Consider the following generic contingency table for a SNP, in a GWAS with $\frac{N}{2}$ cases and $\frac{N}{2}$ controls. The table should be read as follows. There are α cases with SNP value 0 and β cases with value 1. The total number of patients with SNP values 0 and 1 are, respectively, m and n.

SNP value	Cases	Controls
0	α	$m - \alpha$
1	β	$n - \beta$
2	$\frac{N}{2} - \alpha - \beta$	$\frac{N}{2} - m + \alpha - n + \beta$

In a typical GWAS, only SNPs with a MAF larger than some threshold (e.g. 0.05) are considered. Thus, it is reasonable to assume that the **margins** of the contingency table are positive ($m > 0$, $n > 0$, $N - m - n > 0$). Uhler et al. [19] show that the χ^2-statistic of a SNP is then given by

$$\chi^2 = \frac{(2\alpha - m)^2}{m} + \frac{(2\beta - n)^2}{n} + \frac{(2\alpha - m + 2\beta - n)^2}{N - m - n} .$$

Existing Techniques.

Methods for the differentially-private release of SNPs with high χ^2-statistics have been studied by Uhler et al. [19], Johnson and Shmatikov [10], and more recently Yu et al. [22]. When the number of cases and controls are equal, the sensitivity of the χ^2-statistic is $\frac{4N}{N+2}$ [19]. For the general case where the size of the case and control groups are not necessarily equal, the χ^2-statistic and its sensitivity are given in [22]. We consider two exponential mechanisms for outputting M SNPs with high χ^2-statistics and satisfying DP. As noted in [10], the value M of significant SNPs (with a χ^2 score above a given threshold) can also be computed in a differentially private manner. In the following, we assume the total number of SNPs in the study to be M'.

Yu et al. propose a very simple algorithm (Algorithm 1) that directly uses the χ^2-statistics of the SNPs as the score function in the exponential mechanism. Algorithm 1 is ϵ-differentially private [3, 22]. Note that as the number of output SNPs M grows large, the sampling probabilities tend to be uniform. Thus, it is not necessarily beneficial to output more SNPs, in the hope that the SNPs with the highest true statistics will be output.

Algorithm 1 Differentially private release of associated SNPs, using the exponential mechanism [22].

```
Input: The privacy budget $\epsilon$, the sensitivity $s$ of the $\chi^2$-statistic, the number of SNPs $M$ to release.
Output: $M$ SNPs
For $i \in \{1, \dots, M'\}$, compute the score $q_i$ as the $\chi^2$-statistic of the $i^{\text{th}}$ SNP.
Sample $M$ SNPs (without replacement), where SNP $i$ has probability proportional to $\exp\left(\frac{\epsilon \cdot q_i}{2 \cdot M \cdot s}\right)$.
```

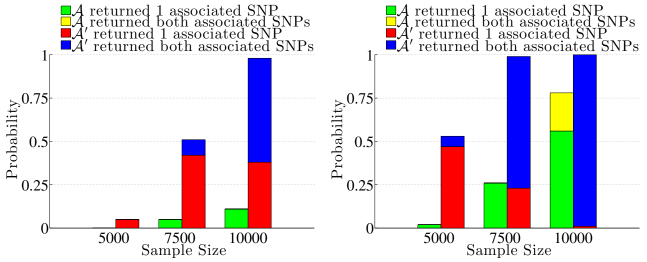
Johnson and Shmatikov [10] propose a general framework that performs multiple queries used in typical GWAS and guarantees differential privacy. They use the exponential mechanism with a specific distance score function. We will focus on their `LocSig` mechanism that outputs M significant SNPs similarly to Algorithm 1. The sole difference is that they use a different score function than the χ^2-statistic.

Let the *distance-to-significance* of a contingency table be defined as the minimal number of SNP values to be modified, in order to obtain a contingency table with a p-value or χ^2-statistic deemed as significant (beyond some pre-defined threshold). Their algorithm for outputting M significant SNPs is then the same as Algorithm 1, where the scores q_i are replaced by the distance-to-significance score, whose sensitivity s can easily be seen to be 1.

As noted by Yu et al. [22], computing these distance scores exactly can be a daunting task for 3×2 contingency tables. They suggest instead to approximate the true distance-to-significance by a greedy approach that only considers edits introducing a maximal change in the χ^2-statistic or p-value. In our experiments, we follow the same approach.

Both of the mechanisms we discussed are subject to a standard tradeoff between privacy, utility and dataset size. We illustrate this tradeoff for Algorithm 1 (see [19] and [22] for details). The tradeoff between privacy and utility is straightforward as the sampling probabilities depend on ϵ. For the dependency on the dataset size, note that by definition, an unassociated SNP is expected to have a χ^2-statistic of 0, regardless of N (this is the null hypothesis). However, if the SNP is correlated to the disease status, we can verify that the value of the χ^2-statistic grows linearly with N. Thus, as N grows, the *gap* between the χ^2-statistics of associated and unassociated SNPs grows as well. Nevertheless, the sensitivity $\Delta\chi^2$ remains bounded above by 4. Combining both observations, we see that the larger N gets, the less probable it is that the algorithm outputs unassociated SNPs. Thus Algorithm 1 achieves high utility for very large datasets.

We show that by considering a weaker but practically significant adversarial model, we require much less patient data in order to achieve high medical utility, thus rendering such privacy protecting mechanisms more attractive and applicable for medical research. Spencer et al. [18] note that a GWAS with 2000 cases and controls necessitates a budget of about \$2,000,000 for standard genotyping chips. Obtaining an acceptable utility-privacy tradeoff even for reasonably large studies is thus an interesting goal.



(a) $\gamma = 1.3$

(b) $\gamma = 1.5$

Figure 5: Utility of mechanisms $\mathcal{A}$ and $\mathcal{A}'$, when outputting 2 SNPs using Algorithm 1 from [22].

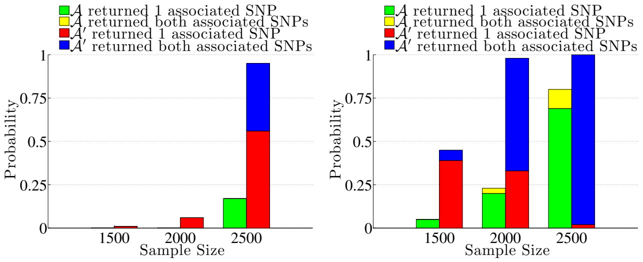
Results.

We evaluate different privacy mechanisms on the GWAS simulations from [19], obtained from the Hap-Sample simulator [21]. The studies consist of 8532 SNPs per participant, typed on chromosomes 9 and 13 using the AFFY 100k array. There are two causative SNPs with an additive effect. We consider mechanisms that use either Algorithm 1 or the `LocSig` mechanism to output 2 SNPs. As a measure of utility, we use the probability (averaged over 1000 runs) that a mechanism outputs either 1 or both of the causative SNPs.

We do not compare the mechanisms from Yu et al. and Johnson and Shmatikov directly (see [22] for a full comparison). Instead, we evaluate how the utility of these mechanisms behave, for a bounded adversarial model close to those models used in the attacks we described in Section 4.2. To this end, we fix a level γ of positive membership-privacy and consider mechanisms that protect against arbitrary priors in $\mathbb{D}_B$ (equivalent to $(\ln \gamma)$-DP) or bounded priors in $\mathbb{D}_B^{0.5}$ (corresponds to a weaker level of DP).

We begin with two privacy mechanisms $\mathcal{A}$ and $\mathcal{A}'$ that use Algorithm 1 to release 2 SNPs and satisfy PMP under $\mathbb{D}_B$ and $\mathbb{D}_B^{0.5}$, respectively. For datasets of sizes $N \in \{5000, 7500, 10000\}$ and PMP parameters $\gamma \in \{1.3, 1.5\}$, we compare the utility of $\mathcal{A}$ and $\mathcal{A}'$, and display the results in Figure 5. We see that for a fixed level of PMP, the bounded adversarial model leads to significantly higher utility. Consider the results for $\gamma = 1.5$. Mechanism $\mathcal{A}$, which satisfies $(1.5, \mathbb{D}_B)$-PMP, requires at least 10000 patients to achieve significant utility. Even in such a large study, the mechanism fails to output any of the causative SNPs in about 25% of the experiments. For $\mathcal{A}'$, which satisfies $(1.5, \mathbb{D}_B^{0.5})$-PMP, we achieve a better utility with only 7500 patients, and quasi-perfect utility for 10000 patients. By focusing on a more reasonable adversarial threat, we thus achieve a good tradeoff between privacy and utility, for much smaller datasets. This is as an attractive feature for medical research, where large patient datasets are typically expensive to obtain.

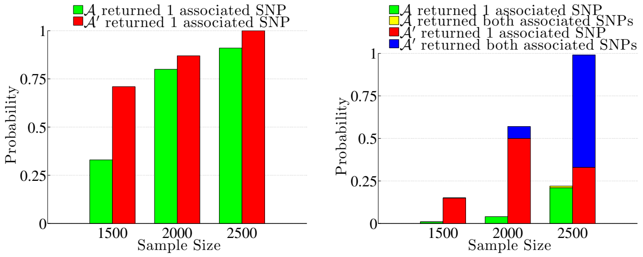
We now consider two privacy mechanisms $\mathcal{A}$ and $\mathcal{A}'$ that use the `LocSig` mechanism to release 2 SNPs. To compute the distance scores, we fix a threshold of 10^{-10} on the p-values, such that exactly 2 SNPs reach this threshold. As before, the mechanisms satisfy positive membership-privacy under $\mathbb{D}_B$ and $\mathbb{D}_B^{0.5}$, respectively. In our particular example, `LocSig` provides better results than Algorithm 1, and we actually achieve similar utility for smaller datasets. For datasets of sizes $N \in \{1500, 2000, 2500\}$ and PMP parameters $\gamma \in \{1.3, 1.5\}$, we compare the utility of $\mathcal{A}$ and $\mathcal{A}'$, and we display the results in Figure 6.



(a) $\gamma = 1.3$

(b) $\gamma = 1.5$

Figure 6: Utility of mechanisms $\mathcal{A}$ and $\mathcal{A}'$, when outputting 2 SNPs using the `LocSig` mechanism from [10].



(a) $M = 1$

(b) $M = 3$

Figure 7: Utility of mechanisms $\mathcal{A}$ and $\mathcal{A}'$, when outputting M SNPs using `LocSig` [10] with $\gamma = 1.5$.

Again, there is a significant improvement in utility if we consider a bounded adversarial model. Although the `LocSig` mechanism yields higher accuracy than the exponential method from Algorithm 1 in this case, we re-emphasize that computing the distance scores has a much higher complexity than the computation of the χ^2-statistics [22]. Deciding upon which method to use in practice is thus subject to a tradeoff between utility and computational cost.

Alternatively, we could consider increasing our utility by releasing $M > 2$ SNPs. However, as the exponential mechanisms we considered associate probabilities proportional to M to each SNP, it is unclear whether we should expect higher utility by increasing M. Obviously, if we were to let M approach the total number of SNPs, the recall would be maximized. Hence, we also consider the precision (ratio of output SNPs that are significant). In Figure 7, we evaluate the utility of `LocSig` with $\gamma = 1.5$, for $M = 1$ and $M = 3$. We see that for $M = 3$, the utility is worse than for $M = 2$, therefore confirming that the increased data perturbation eliminates the potential gain in recall. Also, in this case the precision is naturally upper bounded by $\frac{2}{3}$. An interesting tradeoff is given by selecting $M = 1$. Although recall can not exceed $\frac{1}{2}$, we see that for small datasets ($N \leq 2000$), the utility actually is higher than for $M = 2$.

Finally, we compare the privacy-utility tradeoff for a range of bounds $[a, b]$ on the adversary's prior belief. In Figure 8, we display the probability that Algorithm 1 outputs at least one or both of the causative SNPs in a GWAS with $N = 7500$, while providing PMP with $\gamma = 1.5$. As we can see, even if the considered adversary has only a small degree of *a priori* uncertainty about an individual's presence in the dataset, we still obtain a significant gain in utility compared to the setting where the adversary's prior is unbounded.

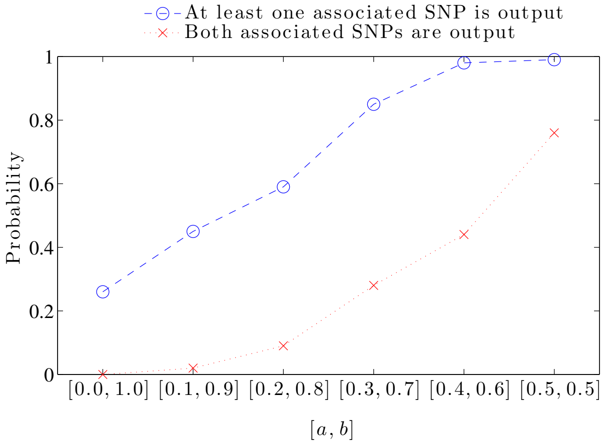


Figure 8: Probability that Algorithm 1 outputs both, or at least one of the causative SNPs, when guaranteeing PMP with $\gamma = 1.5$ against adversaries with prior $\mathbb{D}_B^{[a,b]}$.

Discussion.

For both of the exponential mechanisms we considered, our results show that by focusing on an adversarial setting with bounded prior knowledge, we can attain the same PMP guarantees as for adversaries with arbitrary priors and retain a significantly higher utility. As we argued that the adversarial model with priors in $\mathbb{D}_B^{0.5}$ is relevant in regard to attacks against GWAS, this shows that we can achieve a reasonable level of protection against these attacks and also guarantee an acceptable level of medical utility for datasets smaller (and thus cheaper) than previously reported.

We stress that the applicability of our results need not be limited to GWAS or even to genomic privacy in general. Indeed, we could consider applications in other domains where DP has been proposed as a privacy notion, as a bounded adversarial setting makes sense in many practical scenarios. As we will see in Section 5, our results can also be adapted to cover the case of unbounded-DP, thus further extending their applicability to other use-cases of differential privacy. Examples of settings where DP mechanisms have been proposed, and yet an adversary with incomplete background knowledge appears reasonable, can be found in location privacy [1] or data mining [7] for instance.

In scenarios where DP is applied to protect membership disclosure, we would benefit from considering whether the adversarial setting of DP is reasonable, or whether a bound on an adversary's prior belief is practically significant. Depending on the identified adversaries, we can select an appropriate level of noise to guarantee PMP, according to the model derived in Section 3.

5. THE CASE OF UNBOUNDED-DP

The characterization of unbounded-DP in the PMP framework is a little more subtle than for bounded-DP. Li et al. introduce a uni-directional definition of unbounded-DP.

Definition 9 (Positive Unbounded-DP [14])**.** *A mechanism $\mathcal{A}$ satisfies ϵ-positive unbounded-DP if and only if for any dataset T, any entity t not in T, and any $S \subseteq \mathsf{range}(\mathcal{A})$,*

$$\Pr\left[\mathcal{A}(T \cup \{t\}) \in S\right] \leq e^{\epsilon} \cdot \Pr\left[\mathcal{A}(T) \in S\right] \ . \qquad (14)$$

In this definition, we consider only neighboring datasets obtained by adding a new entity (and not by removing an entity). Note that satisfying ϵ-unbounded DP trivially implies ϵ-positive unbounded-DP.

For this definition, the results we obtained for bounded-DP can be applied rather straightforwardly to (positive) unbounded-DP. Li et al. [14] provide results analogous to Lemma 2 and Theorem 1, by replacing the family $\mathbb{D}_B$, by the family $\mathbb{D}_I$ of mutually-independent distributions.

Lemma 4 ([14])**.** *If $\mathcal{A}$ satisfies ϵ-positive unbounded DP, then for any $\mathcal{D} \in \mathbb{D}_I$ we have $\frac{\Pr[S|t]}{\Pr[S|\neg t]} \leq e^{\epsilon}$.*

Theorem 3 ([14])**.** *A mechanism $\mathcal{A}$ satisfies ϵ-positive unbounded DP if and only if it satisfies $(e^{\epsilon}, \mathbb{D}_I)$-PMP.*

From here on, our analysis from Section 3 can be directly applied to the case of unbounded-DP. We first define a family of bounded prior distributions.

Definition 10 (Restricted MI Distributions)**.** *For $0 < a \leq b < 1$, the family $\mathbb{D}_I^{[a,b]} \subset \mathbb{D}_I$ contains all MI distributions for which $\Pr[t] \in [a, b] \cup \{0, 1\}$, for all entities t. If $a = b$, we simply denote the family as $\mathbb{D}_I^{a}$.*

Finally, we obtain an analogous result to Theorem 2, by characterizing the level of (positive) unbounded-DP that guarantees a level γ of PMP under a restricted MI distribution family.

Theorem 4. *A mechanism $\mathcal{A}$ satisfies $(\gamma, \mathbb{D}_I^{[a,b]})$-PMP, for $0 < a \leq b < 1$, if $\mathcal{A}$ satisfies ϵ-positive unbounded-DP, where*

$$e^{\epsilon} = \begin{cases} \min\left(\frac{(1-a)\gamma}{1-a\gamma}, \ \frac{\gamma+b-1}{b}\right) & \textit{if } a\gamma < 1, \\ \frac{\gamma+b-1}{b} & \textit{otherwise} \ . \end{cases}$$

6. CONCLUSION AND FUTURE WORK

We have investigated possible relaxations of the adversarial model of differential privacy, the strength of which has been questioned by recent works. By considering the problem of protecting against set membership disclosure, we have provided a complete characterization of the relationship between DP and PMP for adversaries with limited prior knowledge. We have argued about the practical significance of these weaker adversarial settings and have shown that we can achieve a significantly higher utility when protecting against such bounded adversaries.

We have proposed a simple model for the selection of the DP parameter, that consists in identifying a practically significant adversarial setting, as well as an appropriate bound on an adversary's posterior belief. We have illustrated these points with a specific example on genome-wide association studies and have shown that privacy threats identified in the literature can be re-cast into our bounded adversarial model, which leads to a better tradeoff between privacy guarantees and medical utility. Evaluating the applicability of our model to other privacy domains, as well as the corresponding utility gain, is an interesting direction for future work.

Our results from Theorems 1 and 4 show that when we consider an adversary with limited prior knowledge, satisfying DP provides a *sufficient* condition for satisfying PMP. An interesting direction for future work is to investigate whether PMP under distribution families $\mathbb{D}_B^{[a,b]}$ and $\mathbb{D}_I^{[a,b]}$

can be attained by other means than through DP. For instance, in their work on membership privacy, Li et al. propose a simple mechanism for outputting the maximum of a set of values, that satisfies PMP for the family $\mathbb{D}_I^{0.5}$ but does not satisfy any level of DP [14]. It is unknown whether similar mechanisms could be designed for other queries (such as those we considered in our GWAS scenario), in order to potentially improve upon the privacy-utility tradeoff of DP.

Acknowledgments

We thank Mathias Humbert and Huang Lin for helpful comments.

7. REFERENCES

[1] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, CCS '13, pages 901–914, New York, NY, USA, 2013. ACM.

[2] R. Bassily, A. Groce, J. Katz, and A. Smith. Coupled-worlds privacy: Exploiting adversarial uncertainty in statistical data privacy. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 439–448. IEEE, 2013.

[3] R. Bhaskar, S. Laxman, A. Smith, and A. Thakurta. Discovering frequent patterns in sensitive data. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 503–512. ACM, 2010.

[4] C. Dwork. Differential privacy. In *Automata, languages and programming*, pages 1–12. Springer, 2006.

[5] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Conference on Theory of Cryptography*, TCC'06, pages 265–284, Berlin, Heidelberg, 2006. Springer-Verlag.

[6] M. Fredrikson, E. Lantz, S. Jha, S. Lin, D. Page, and T. Ristenpart. Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 17–32, San Diego, CA, Aug. 2014. USENIX Association.

[7] A. Friedman and A. Schuster. Data mining with differential privacy. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 493–502. ACM, 2010.

[8] J. Gehrke, M. Hay, E. Lui, and R. Pass. Crowd-blending privacy. In *Advances in Cryptology–CRYPTO 2012*, pages 479–496. Springer, 2012.

[9] N. Homer, S. Szelinger, M. Redman, D. Duggan, W. Tembe, J. Muehling, J. V. Pearson, D. A. Stephan, S. F. Nelson, and D. W. Craig. Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLoS genetics*, 4(8):e1000167, 2008.

[10] A. Johnson and V. Shmatikov. Privacy-preserving data exploration in genome-wide association studies. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 1079–1087, New York, NY, USA, 2013. ACM.

[11] D. Kifer and A. Machanavajjhala. No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, SIGMOD '11, pages 193–204, New York, NY, USA, 2011. ACM.

[12] J. Lee and C. Clifton. Differential identifiability. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, pages 1041–1049, New York, NY, USA, 2012. ACM.

[13] N. Li, W. Qardaji, and D. Su. On sampling, anonymization, and differential privacy or, k-anonymization meets differential privacy. In *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security*, ASIACCS '12, pages 32–33, New York, NY, USA, 2012. ACM.

[14] N. Li, W. Qardaji, D. Su, Y. Wu, and W. Yang. Membership privacy: a unifying framework for privacy definitions. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, CCS '13, pages 889–900, New York, NY, USA, 2013. ACM.

[15] E. Lui and R. Pass. Outlier privacy. In Y. Dodis and J. Nielsen, editors, *Theory of Cryptography*, volume 9015 of *Lecture Notes in Computer Science*, pages 277–305. Springer Berlin Heidelberg, 2015.

[16] A. Machanavajjhala, J. Gehrke, and M. Götz. Data publishing against realistic adversaries. *Proc. VLDB Endow.*, 2(1):790–801, Aug. 2009.

[17] F. McSherry and K. Talwar. Mechanism design via differential privacy. In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*, pages 94–103. IEEE, 2007.

[18] C. C. Spencer, Z. Su, P. Donnelly, and J. Marchini. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS genetics*, 5(5):e1000477, 2009.

[19] C. Uhler, A. Slavkovic, and S. E. Fienberg. Privacy-preserving data sharing for genome-wide association studies. *Journal of Privacy and Confidentiality*, 5(1), 2013.

[20] R. Wang, Y. F. Li, X. Wang, H. Tang, and X. Zhou. Learning your identity and disease from research papers: Information leaks in genome wide association study. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*, CCS '09, pages 534–544, New York, NY, USA, 2009. ACM.

[21] F. A. Wright, H. Huang, X. Guan, K. Gamiel, C. Jeffries, W. T. Barry, F. P.-M. de Villena, P. F. Sullivan, K. C. Wilhelmsen, and F. Zou. Simulating association studies: a data-based resampling method for candidate regions or whole genome scans. *Bioinformatics*, 23(19):2581–2588, 2007.

[22] F. Yu, S. E. Fienberg, A. B. Slavković, and C. Uhler. Scalable privacy-preserving data sharing methodology for genome-wide association studies. *Journal of biomedical informatics*, 2014.