Estimating Group Means Under Local Differential Privacy

René Raab
Friedrich-Alexander-Universität
Erlangen-Nürnberg, Germany

Arijana Bohr
Friedrich-Alexander-Universität
Erlangen-Nürnberg, Germany

Kai Klede
Friedrich-Alexander-Universität
Erlangen-Nürnberg, Germany

Benjamin Gmeiner
Novartis Pharma GmbH
Nürnberg, Germany

Bjoern M. Eskofier
Friedrich-Alexander-Universität
Erlangen-Nürnberg, Germany
Institute of AI for Health, Helmholtz
Zentrum München - German
Research Center for Environmental
Health
Neuherberg, Germany

Abstract

The European Health Data Space (EHDS) aims to enable the sharing of health data across Europe to improve healthcare and research. While the EHDS mandates anonymization or pseudonymization of shared health data, these techniques may still allow adversaries to re-identify individuals. Local differential privacy (LDP) has been proposed as a formal privacy guarantee that can help mitigate this issue. In this paper, we consider a common problem when analyzing health data: estimating means for different groups. We discuss a generic privacy-preserving method for approximating the means of different groups in a decentralized setting where both the group and the value are considered private. We show that four concrete instantiations of the method based on existing mean estimation methods (Laplace, Bernoulli, Piecewise, and NPRR) are locally differentially private. We evaluate their performance on synthetic and real-world medical datasets. Our results show that the proposed methods can accurately estimate the group means, while maintaining privacy. However, similar to other LDP algorithms, our approach requires a sufficient amount of data (in our case a sufficient amount of samples per group) combined with a sufficiently large privacy budget ε to produce accurate results. We discuss concrete practical issues like choosing an appropriate input range, dealing with large privacy budgets through the use of the shuffle model of differential privacy, and the need for further analysis techniques to make LDP solutions applicable to practical medical data analysis.

Keywords

local differential privacy, data analysis, group means, decentralized data, mean estimation

1 Introduction

Trust plays a vital role in medical research as participants share sensitive personal health information with research organizations. This trust relationship is typically acceptable for individuals in the context of reputable institutions conducting limited studies, however, the picture is shifting dramatically with initiatives like the European Union's upcoming legislation on the European Health Data Space (EHDS) [21]. This legislation aims to make electronic health data available to a broader range of organizations and research institutions, potentially without the explicit consent of the individuals concerned. By expanding the scale of health data sharing and the number of people that have access to shared data, the EHDS increases the likelihood for the potential exposure of personal health information [40]. Although the EHDS requires the anonymization or pseudonymization of shared health data, significant vulnerabilities remain and may allow the re-identification of individuals that are part of the underlying private dataset [5, 41].

Local Differential Privacy (LDP) provides a formal privacy guarantee and has the potential to mitigate these problems. Under LDP, a certain amount of noise is added to each individual data record before it is shared with a data collector. This means that the data collector never sees the raw data, and even if they are compromised, they can only learn a certain amount of information about the data.

In this paper, we consider a common problem when analyzing health data: estimating the mean of a continuous numerical value for different groups. More specifically, we consider the scenario where n participants each own a private value (e.g., a survey response, a lab measurement, or other sensitive information) and belong to a private group (e.g., a demographic attribute, a diagnosis, or other sensitive categories). The goal is to estimate the mean of the private values for each group while preserving the privacy of the individual values and group memberships using LDP. This allows raw health data to remain with its original providers while still enabling statistical analysis across institutions or individuals.

Related Work. The main building blocks of our methods are existing LDP methods for mean estimation and frequency estimation. Estimation of the mean of a scalar value is a common problem in the LDP literature and has been discussed in many articles [10, 12–14, 24, 30, 38, 42, 46]. Frequency estimation, i.e., determining how often a discrete value appears in private data, has gotten even more attention in the literature [1, 4, 9, 14, 20, 32–34, 37, 38, 42, 43, 48].

However, few works discuss the goal of this paper, i.e., the mean estimation for different groups. Ding et al. [11] introduce a method for performing hypothesis tests to compare the mean between two

Proceedings on Privacy Enhancing Technologies 2025(4), 236–274

https://doi.org/10.56553/popets-2025-0129

groups, but primarily focus on the application of A/B testing and therefore do not consider the group as private. Friedberg and Rogers [23] present a method to estimate the mean of a continuous value for different groups, but only consider the group information as private, not the value. Juarez and Korolova [31] consider both the group and the value as private and introduce two methods (Laplace and Bernoulli) for estimating the mean for the different groups.

Contributions. We extend the work of Juarez and Korolova [31] in several ways: (1) We generalize their method, introduce two new variants (Group Piecewise and Group NPRR), and show that they are ε-LDP. (2) We provide a corrected privacy guarantee for the Bernoulli method (Theorem 2). (3) We show that the modification to the Laplace method proposed by Juarez and Korolova [31] to use different noise scales inside the method does not provide LDP (Proposition 2). (4) We empirically compare the new and existing methods on synthetic and real-world datasets. (5) We evaluate the optimal parameter setting for the new Group Piecewise and Group NPRR methods.

Organization. The remainder of this paper is organized as follows: In Section 2, we introduce the definition of LDP. Section 3 formally defines the problem of estimating the mean of a private value for different groups and introduces relevant LDP methods used as building blocks for our methods. In Section 4, we introduce our generic method for estimating the means of different groups and discuss four concrete instantiations, their privacy guarantees, and corresponding estimators. Section 5 describes the experimental setup and presents the results of our experiments on synthetic and real-world medical datasets. In Section 6, we discuss the limitations of the methods and discuss open problems and potential solutions. Finally, Section 7 concludes the paper and provides an outlook on future work. The proofs for all theorems and propositions are provided in Section D in the appendix.

2 Local Differential Privacy

Differential privacy (DP) exists in several variants, each tailored to different trust models and privacy requirements [17]. These models offer distinct trade-offs between privacy and utility, depending on the level of trust placed in different entities involved in the data processing pipeline.

Central DP, first introduced by Dwork et al. [16], assumes a trusted central entity that collects private data from participants and applies a DP mechanism to produce privacy-preserving aggregate results. This model relies on participants' trust that the central entity will not misuse or leak their private data. It is useful in scenarios where data is collected centrally and statistics should be released in a privacy-preserving manner.

However, this trust assumption is not suitable for all applications, especially when participants are unwilling to share their data with a central entity. LDP addresses this issue by eliminating the need for a trusted curator. Instead of applying privacy mechanisms to a centralized database, LDP operates directly on individual data points, making it the primary focus of this paper. Each participant randomizes their private data *before* submitting it to an aggregator who can then analyze the data or compute approximate aggregate statistics. The perturbation is chosen carefully to enable the estimation of aggregate statistics by the analyst while hiding individual private data points.

LDP was first formalized by Kasiviswanathan et al. [35] and is based on a local randomizer which performs the randomization on the participant's side. We define LDP as follows:

Definition 1 (Local Differential Privacy). *A randomized algorithm M with domain D is* ε-locally differentially private *(an ε-DP local randomizer; $\varepsilon > 0$) if for all $\mathcal{S} \subseteq \text{Range}(M)$ and for all pairs of a participant's values $x, y \in D$:*

$$\Pr[M(x) \in \mathcal{S}] \leq \exp(\varepsilon) \Pr[M(y) \in \mathcal{S}].$$

While LDP removes the need for a trusted curator, it introduces new challenges. Because noise must be added locally, the overall noise in the data increases, often reducing the utility of the results. To ensure accurate results, LDP mechanisms often require either a large number of participants or the choice of a larger privacy budget ε.

3 Problem Definition

We consider each participant $i \in \{1, 2, \ldots, n\}$ out of n total participants to have private data (g_i, v_i), where $g_i \in G, |G| = d$ is indicating the group membership and $v_i \in D$ is the private value of interest. Without loss of generality, we assume that $D = [-1, 1]$ for continuous values.[1] With n_g, we denote the number of participants belonging to group g, i.e., $n_g = \sum_i^n \mathbf{1}_{g_i=g}$. We use $\mathbf{1}_A$ to denote the indicator function which equals 1 if A is true and 0 otherwise. To ensure the privacy guarantee, each participant has to apply an LDP mechanism M to their private value and group membership before sending the perturbed data $(g_i', v_i') = M(g_i, v_i)$ to the aggregator. The aggregator then uses the perturbed data to estimate the true group means $m_g = \frac{1}{n_g} \sum_i^n \mathbf{1}_{g_i=g} \cdot v_i$ from the perturbed data (g_i', v_i').

3.1 (Generalized) Randomized Response

The core building block of our LDP mechanisms is the generalized randomized response mechanism. Randomized response (RR) was first introduced by Warner [45] as a method to collect sensitive information while preserving the privacy of the participants. The idea is that participants do not answer a sensitive question truthfully, but instead randomize their response before submitting it. The randomized response mechanism is a simple way to achieve LDP. More formally, the randomized response mechanism M_{RR} for a binary input $x \in \{0, 1\}$ is defined by the probability [44]

$$\Pr[M_{\text{RR}}(x, \varepsilon) = y] = \begin{cases} \frac{e^\varepsilon}{e^\varepsilon + 1} & \text{if } x = y \\ \frac{1}{e^\varepsilon + 1} & \text{if } x \neq y \end{cases}. \tag{1}$$

Using this definition, M_{RR} is ε-LDP [44].

The randomized response mechanism can be generalized to non-binary inputs. The generalized randomized response (GRR) mechanism M_{GRR} for a discrete input domain X of size $|X| = d$ is defined by the probability [33, 44]

$$\Pr[M_{\text{GRR}}(x, X, \varepsilon) = y] = \begin{cases} \frac{e^\varepsilon}{e^\varepsilon + d - 1} & \text{if } x = y \\ \frac{1}{e^\varepsilon + d - 1} & \text{if } x \neq y \end{cases}, \tag{2}$$

[1] Any bounded interval can be transformed to this interval. Section C in the appendix discusses the concrete transform used for the experiments in this paper.

where $x, y \in X$. This definition of GRR is ε-LDP [44] and reduces to RR for $d = 2$. Throughout this paper, we use $M_{\text{GRR}}(x, d, \varepsilon)$ as a shorthand for $M_{\text{GRR}}(x, \{1, 2, \dots, d\}, \varepsilon)$.

3.2 Mean Estimation Mechanisms

Several mechanisms have been proposed for locally differentially private mean estimation (without group information). We discuss the most relevant mechanisms in the following, before extending them to group-wise mean estimation in the next section. We refer the reader to Raab et al. [39] for a comprehensive overview of the state-of-the-art in LDP mechanisms and to the original papers for more details.

3.2.1 Laplace Mechanism. The Laplace mechanism is one of the first and most well-known LDP mechanisms. It was first introduced by Dwork et al. [16] for DP and Kasiviswanathan et al. [35] for LDP and is based on additive noise drawn from the Laplace distribution. We define the Laplace mechanism as follows:

Definition 2 (Laplace mechanism). *Given a private value $v \in [-1, 1]$ and privacy budget $\varepsilon > 0$, the Laplace mechanism M_{Lap} is defined as*

$$M_{\text{Lap}}(v, \varepsilon) = v + \text{Lap}\left(\frac{2}{\varepsilon}\right), \tag{3}$$

where $\text{Lap}(\lambda)$ denotes a sample from the Laplace distribution with probability density function $f_\lambda(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$.

3.2.2 Bernoulli-based Mechanisms. Bernoulli-based mechanisms have been introduced by several works independently [10, 13, 38], but are in fact equivalent [39]. Their main idea is to use a Bernoulli random variable to discretize the input value before applying the randomized response mechanism to the resulting binary value. We give a general definition in Algorithm 1.

Algorithm 1 Generic Bernoulli-Based Mechanism M_{Bern}

Input: Client's value $v \in [-1, 1]$, privacy budget $\varepsilon > 0$
Output: Perturbed value $v' \in \{-1, 1\}$

$B \sim \text{Bernoulli}\left(\frac{1+v}{2}\right)$
$v' \leftarrow 2M_{\text{RR}}(B, \varepsilon)$ - 1
return v'

3.2.3 Piecewise Mechanism. The piecewise mechanism was first introduced by Wang et al. [42] and is based on the idea of partitioning the output space into three intervals with the "middle" interval being centered around the true value. The response is then created by sampling from the output space, where sampling from the "middle" interval is more likely than sampling from the "outer" intervals. The ε-LDP piecewise mechanism is defined as follows:

Definition 3 (Piecewise Mechanism [42]). *Given a private value $v \in [-1, 1]$ and privacy budget $\varepsilon > 0$, the piecewise mechanism M_{PW} outputs a perturbed value $v' \in [-C, C]$ where $C = \frac{\exp(\varepsilon/2)+1}{\exp(\varepsilon/2)-1}$. The mechanism is defined by the probability*

$$\Pr[M_{\text{PW}}(v, \varepsilon) = x] = \begin{cases} p & \text{if } x \in [l(v), r(v)] \\ \frac{p}{\exp(\varepsilon)} & \text{if } x \in [-C, l(v)) \cup (r(v), C] \end{cases}, \tag{4}$$

where $l(v) = \frac{C+1}{2}v - \frac{C-1}{2}$, $r(v) = l(v) + C - 1$, and $p = \frac{\exp(\varepsilon) - \exp(\varepsilon/2)}{2\exp(\varepsilon/2)+2}$.

3.2.4 NPRR Mechanism. Non-Parametric Randomized Response (NPRR) was introduced by Waudby-Smith et al. [46] as a generalization of the binary Bernoulli-based mechanisms. While their general solution is interactive and allows for individual privacy budgets for each participant, we use a non-interactive simplification of the NPRR mechanism with a single privacy budget for all participants. Algorithm 2 gives the simplified NPRR mechanism (see Section B in the appendix for a discussion of the simplification). While the original algorithm is defined for inputs in $[0, 1]$, the simplified algorithm works with inputs in $[-1, 1]$ to better align with the other methods in this paper.

Algorithm 2 Simplified Non-Parametric Randomized Response (NPRR) Mechanism M_{NPRR}

Input: Client's value $v \in [-1, 1]$, privacy budget $\varepsilon > 0$, discretization parameter $k \in \{1, 2, \dots\}$
Output: Perturbed value $v' \in \{-1, \frac{2-k}{k}, \frac{4-k}{k} \dots, \frac{2k-k}{k} = 1\}$

$v^f \leftarrow \lfloor \frac{v+1}{2} \cdot k \rfloor / k$
$B \sim \text{Bernoulli}\left(k(\frac{v+1}{2} - v^f)\right)$
$y \leftarrow v^f + B/k$
$v' \leftarrow 2M_{\text{GRR}}(y, \{0/k, 1/k, \dots, k/k\}, \varepsilon) - 1$
return v'

4 Methods

Juarez and Korolova [31] first discussed the idea of estimating the mean of a group with local differential privacy with the aim to estimate the performance of a federated learning algorithm for different demographic groups. We generalize their approach to work with any mechanism for mean estimation and discuss four concrete instantiations based on the mechanisms introduced in the previous section.

The generic mechanism for perturbing group and value is given in Algorithm 3 and works as follows. Each client applies GRR to perturb their group g as $g' = M_{\text{GRR}}(g, d, \varepsilon_1)$, where $|G| = d$ is the number of groups. If GRR changes the group ($g' \neq g$), the client replaces their true value v by a different "neutral" value e_0 – in the case of the four specific algorithms we discuss in this paper, we set $e_0 = 0$. Other values are possible, but potentially lead to more complex formulas for the resulting mean estimators. The client then perturbs their value v as $v' = M(v, \varepsilon_2)$ using a method-specific mechanism. Note that this means that the client never directly reports the unperturbed "neutral" value e_0 to the aggregator (doing this could leak the fact that $g' \neq g$). Finally, the client sends (g', v') to the aggregator, who can then estimate the mean for each group. The range of v and v' depends on the choice of the mechanism M and is discussed in detail in the following sections.

Algorithm 3 Generic mechanism for group-value responses

Parameters: Mechanism $M : D \to O$, neutral value e_0
Input: Client's group $g \in \{1, 2, \ldots, d\}$, value $v \in D$ and privacy budgets $\varepsilon_1, \varepsilon_2 \in [0, +\infty]$
Output: Perturbed tuple (g', v'), $g' \in \{1, 2, \ldots, d\}$, $v' \in O$

$g' \leftarrow M_{\text{GRR}}(g, d, \varepsilon_1)$
if $g' \neq g$:
  $v \leftarrow e_0$
$v' \leftarrow M(v, \varepsilon_2)$
return (g', v')

Unlike the mechanisms introduced in the previous section, this algorithm takes two parameters ε_1 and ε_2 which influence the overall ε-LDP guarantee of the algorithm. We discuss concrete interactions between $\varepsilon_1, \varepsilon_2$, and ε in the following sections. From the general composition theorem of differential privacy [15, 17], we obtain that the sequential application of an ε_1-LDP mechanism and an ε_2-LDP mechanism is ε-LDP with $\varepsilon = \varepsilon_1 + \varepsilon_2$. In the following sections, we will show that in three out of four cases, we get a better privacy guarantee for the resulting algorithm (i.e., $\varepsilon < \varepsilon_1 + \varepsilon_2$).

To estimate the mean per group, we need an estimate of the number of participants in each group since the true number of participants per group requires knowledge of the private group memberships g_i and is therefore unknown[2]. Since the group information only depends on the output of the GRR mechanism, we can define a general estimator for the number of participants in each group which works regardless of the choice of mechanism M. This is essentially the same estimator as for frequency estimation with GRR [44].

Definition 4 (Estimator of n_g). *Let n_g be the true number of participants in group g, i.e., $n_g = \sum_{i=1}^{n} \mathbf{1}_{g_i=g}$. The estimator $\hat{n}_g$ for the number of participants in group g is given by*

$$\hat{n}_g = \frac{1}{p-q}\left(\sum_{i=1}^{n} \mathbf{1}_{g_i'=g} - nq\right), \tag{5}$$

where $p = \frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}$ and $q = \frac{1-p}{d-1} = \frac{1}{e^{\varepsilon_1}+d-1}$.

Proposition 1. *The estimator $\hat{n}_g$ is unbiased.*

To estimate each group's mean value, we combine the group size estimator $\hat{n}_g$ with a separate estimator for the sum of the values within each group, denoted $\hat{s}_g$. In the following, we introduce four instantiations of Algorithm 3 and corresponding estimators for $\hat{s}_g$.

4.1 Laplace Mechanism

The first mechanism M^G_{Lap} uses the Laplace mechanism $M_{\text{Lap}} : [-1, 1] \to \mathbb{R}$ with $e_0 = 0$ as the neutral value. Following the general Laplace mechanism (Definition 2), the value is perturbed by adding Laplace noise scaled by $\lambda = \frac{2}{\varepsilon_2}$. This method was first introduced by Juarez and Korolova [31], who showed the privacy guarantee which we repeat in Theorem 1.

Theorem 1 (Juarez and Korolova [31]). *The mechanism M^G_{Lap} is ε-LDP with $\varepsilon = \max\left\{\varepsilon_2, \frac{\varepsilon_2}{2} + \varepsilon_1\right\}$.*

[2] Juarez and Korolova [31] assume that the number of participants per group is known.

The optimal privacy allocation for the Laplace mechanism would therefore be $\varepsilon_1 = \frac{\varepsilon}{2}$ and $\varepsilon_2 = \varepsilon$.

Juarez and Korolova [31] also claimed that it would be beneficial to use a different scale for the Laplace noise in the case when the group was changed ($g' \neq g$) by the algorithm. In this variant, the value would be perturbed as before if the group was not changed, but the neutral value e_0 would be perturbed with a different scale for the Laplace noise. However, we show that the resulting mechanism would not satisfy LDP:

Proposition 2. *The mechanism M^G_{Lap} is not ε-LDP if the noise scale in the case $g' = g$ differs from the case $g' \neq g$.*

We now give an estimator for the sum of values within each group. This is based on the estimator and proof given by Juarez and Korolova [31]. However, their original paper provides an estimator for the mean which requires the (unknown) group size to be known.

Definition 5 (Estimator of s_g for the Laplace Mechanism). *Let s_g be the true sum of values in group g and $(g_i', v_i') = M^G_{\text{Lap}}(g_i, v_i)$. The estimator $\hat{s}_g^{\text{Lap}}$ for s_g is given by*

$$\hat{s}_g^{\text{Lap}} = \frac{1}{a}\sum_{i=1}^{n} \mathbf{1}_{g_i'=g} \cdot v_i', \tag{6}$$

where $a = \frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}$.

Proposition 3 (Modified from Juarez and Korolova [31]). *The estimator $\hat{s}_g^{\text{Lap}}$ is unbiased.*

4.2 Binary Bernoulli Mechanisms

The second mechanism M^G_{Bern} uses the Bernoulli mechanism $M_{\text{Bern}} : [-1, 1] \to \{-1, 1\}$ with $e_0 = 0$ as the neutral value. This method was also discussed by Juarez and Korolova [31], who claimed an incorrect privacy guarantee ($\varepsilon_1 = \varepsilon_2 = \varepsilon$) and only considered the case $d = 2$. We extend their proof to general d and provide a corrected privacy guarantee. This method first perturbs the group using GRR and then perturbs the true value (or 0 if the group was changed) using a binary Bernoulli mechanism.

Theorem 2 (Modified from Juarez and Korolova [31]). *The mechanism M^G_{Bern} is ε-LDP with $\varepsilon = \max\{\varepsilon_1 + \ln\left(\frac{2e^{\varepsilon_2}}{e^{\varepsilon_2}+1}\right), \varepsilon_2\}$.*

For a given ε, the optimal privacy allocation would therefore be $\varepsilon_1 = \varepsilon - \ln\left(\frac{2e^{\varepsilon}}{e^{\varepsilon}+1}\right)$ and $\varepsilon_2 = \varepsilon$.

Definition 6 (Estimator of s_g for the Bernoulli mechanism). *Let s_g be the true sum of values in group g and $(g_i', v_i') = M^G_{\text{Bern}}(g_i, v_i)$. The estimator $\hat{s}_g^{\text{Bern}}$ for s_g is given by*

$$\hat{s}_g^{\text{Bern}} = \frac{1}{a(2b-1)}\sum_{i=1}^{n} \mathbf{1}_{g_i'=g} \cdot v_i' \tag{7}$$

where $a = \frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}$ and $b = \frac{e^{\varepsilon_2}}{e^{\varepsilon_2}+1}$.

Proposition 4 (Modified from Juarez and Korolova [31]). *The estimator $\hat{s}_g^{\text{Bern}}$ is unbiased.*

4.3 NPRR Mechanism

The third mechanism M^G_{NPRR} uses the NPRR mechanism $M_{\text{NPRR}} : [-1, 1] \to \{-1, \frac{2-k}{k}, \frac{4-k}{k} \ldots, \frac{2k-k}{k}\}$ with $e_0 = 0$ as the neutral value.

Theorem 3. *The resulting mechanism M^G_{NPRR} is ε-LDP with $\varepsilon = \max\left\{\varepsilon_1 + \ln\left(\frac{(k+1)e^{\varepsilon_2}}{e^{\varepsilon_2}+k}\right), \varepsilon_2\right\}$.*

For a given ε, the optimal privacy allocation would therefore be $\varepsilon_1 = \varepsilon - \ln\left(\frac{(k+1)e^{\varepsilon}}{e^{\varepsilon}+k}\right)$ and $\varepsilon_2 = \varepsilon$.

Definition 7 (Estimator of s_g for the NPRR Mechanism). *Let s_g be the true sum of values in group g and $(g'_i, v'_i) = M^G_{\text{NPRR}}(g_i, v_i)$. The estimator $\hat{s}^{\text{NPRR}}_g$ for s_g is given by*

$$\hat{s}^{\text{NPRR}}_g = \frac{1}{ab} \sum_{i=1}^{n} \mathbf{1}_{g'_i=g} \cdot v'_i, \tag{8}$$

where $a = \frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}$, and $b = \frac{e^{\varepsilon_2}-1}{e^{\varepsilon_2}+k}$.

Proposition 5. *The estimator $\hat{s}^{\text{NPRR}}_g$ is unbiased.*

4.4 Piecewise Mechanism

The final mechanism M^G_{PW} uses the piecewise mechanism $M_{\text{PW}} : [-1, 1] \to [-C, C]$ with $e_0 = 0$ as the neutral value.

Proposition 6. *The mechanism M^G_{PW} is ε-LDP with $\varepsilon = \varepsilon_1 + \varepsilon_2$.*

In the case of the piecewise mechanism, we cannot find a better privacy bound than given by the general composition theorem. Therefore, we cannot derive a clear optimal privacy allocation between ε_1 and ε_2 from this result. In the experiments in this paper, we set $\varepsilon_1 = \varepsilon_2 = \frac{\varepsilon}{2}$.

Definition 8 (Estimator of s_g for the Piecewise Mechanism). *Let s_g be the true sum of values in group g and $(g'_i, v'_i) = M^G_{\text{PW}}(g_i, v_i)$. The estimator $\hat{s}^{\text{PW}}_g$ for s_g is given by*

$$\hat{s}^{\text{PW}}_g = \frac{1}{a} \sum_{i=1}^{n} \mathbf{1}_{g'_i=g} \cdot v'_i, \tag{9}$$

where $a = \frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}$.

Proposition 7. *The estimator $\hat{s}^{\text{PW}}_g$ is unbiased.*

5 Experiments

In this section, we empirically test the performance of the four algorithms by comparing them on synthetic data. We also assess their impact on medical data analysis by inspecting individual use cases from real-world medical datasets.

5.1 Data

First, we use synthetic data as a preliminary validation to verify the correctness of the methods, to find the optimal discretization parameter for NPRR, to compare different settings of ε_1 and ε_2 for the Piecewise method, to compare the performance of the methods in general, and to investigate the impact of unbalanced data. Next, we transition to the AQUILA dataset to represent typical observational studies. Finally, we test the methods on scenarios from the comprehensive MIMIC-IV dataset to align more closely with aspects of secondary use expected in the forthcoming European Health Data Space (i.e., the large-scale sharing of health data for research and development purposes).

As real data does not always conform to the input range $[-1, 1]$ required by the algorithms, we transform the raw data before executing the algorithm and also transform the resulting means back to the correct value range. We describe this process in Section C in the appendix.

5.1.1 Synthetic Data. We use four synthetic datasets to compare the utility of the proposed methods. Each dataset consists of d groups, each with an equal number of participants. The input range for these datasets is $[s, r] = [-1, 1]$ if not stated otherwise in the following sections.

The datasets are defined as follows:

- **Uniform**: This dataset consists of d groups, each with data uniformly sampled from the full domain $[s, r]$. Consequently, all groups have (roughly) the same mean.
- **Normal**: This dataset consists of d groups, with evenly spaced population means μ_g. Each group g has values sampled from a normal distribution $\mathcal{N}\left(\mu_g, \frac{s+r}{5d}\right)$.
- **Constant**: This dataset consists of d groups with the same evenly spaced means as the Normal dataset. However, each group only has values exactly equal to μ_g.
- **Extremum**: This dataset consists of d groups with the same evenly spaced means as the Normal dataset. The values are either s or r, sampled from a Bernoulli distribution to approximate the target mean.

The data distributions for the synthetic datasets are visualized in Figure 21 in the appendix.

5.1.2 AQUILA. To represent data from observational studies, we use the AQUILA dataset [36]. AQUILA is an ongoing, multicenter, prospective, non-interventional study to assess different aspects of treatment with Secukinumab (a monoclonal antibody) in patients with ankylosing spondylitis (AS) and psoriatic arthritis (PsA).

The dataset contains entries for 1912 patients (993 female, 919 male), with 668 patients with AS and 1245 patients with PsA. We provide an overview of the group and value attributes in Table 1. The data distribution for the AQUILA dataset is visualized in Figure 22 in the appendix.

Table 1: Overview of the AQUILA datasets used in the experiments.

(a) Values

Attribute	Input Range
Age (years)	$18 - 100$
BMI (kg m^{-2})	$10 - 60$
Height (m)	$1.0 - 2.3$
Weight (kg)	$30 - 200$

(b) Groups

Attribute	Group Size
Disease	2 (AS, PsA)
Gender	2 (female, male)

5.1.3 MIMIC-IV. We simulate a potential future secondary use in the EHDS through the use of the MIMIC-IV dataset [28]. This dataset provides a large collection of health records collected in the Beth Deaconess Medical Center between 2008 and 2019. We use

MIMIC-IV [26] version 2.2 and pre-process it using the MIMIC code repository [27, 29]. The dataset is provided through PhysioNet [25]. This dataset is a good representation of the anticipated secondary data use in the EHDS context, as it contains routine health care data from a real hospital and has sufficient size.

We only use the hospital module from the MIMIC-IV dataset, which contains data for 431 231 hospital admissions for 299 712 patients and primarily provides categorical properties and few numerical ones. For experiments with this dataset, we consider each individual admission as an individual "participant" in the LDP process. MIMIC-IV only contains data for patients 18 years or older. Table 2 provides a summary of the categories we use for groups and values. The data distributions for MIMIC-IV datasets with a group size of two is visualized in Figure 22 in the appendix.

Table 2: Overview of the MIMIC-IV datasets used in the experiments. "Deceased" is a binary value indicating whether the MIMIC-IV dataset contains information about a patient's death. The values for the admission type and location are given in Table 7 in the appendix.

(a) Values

Attribute	Input Range
Age (years)	18 – 100

(b) Groups

Attribute	Group Size
Gender	2 (female, male)
Admission Type	9 (see Table 7)
Admission Location	11 (see Table 7)
Deceased	2 (no, yes)

5.2 Validation on Synthetic Data

In this section, we evaluate the performance of the proposed methods using the synthetic datasets. We mainly focus on the scaled mean absolute error, which is computed by dividing the absolute error between the estimated and true means by the size of the input range. This scaled error can be interpreted as a percentage relative to the input range, allowing for comparisons across different input ranges. For the group error ratio, we determine the proportion of incorrect group responses ($g_i' \neq g_i$) relative to the total number of participants n.

5.2.1 Choosing the Discretization Parameter for NPRR. Before comparing the performance of the four group mean estimation methods, we first investigate the influence of the discretization parameter k on the NPRR mechanism M^G_{NPRR}. A larger k should reduce the error introduced by the randomized rounding (discretization) procedure, however it also increases the amount of items for the response and could therefore increase the error introduced by GRR. We investigate this influence in this first experiment by executing the NPRR mechanism for $k \in \{1, 2, 3, 4, 8, 16, 32\}$ on data with differing amounts of groups and group sizes. For this experiment, we use synthetic data with equal group sizes, where the group size is set to 10 000 participants and the number of groups is varied between 2, 8, and 64. Each simulation run was repeated 200 times to account for the randomness in the mechanism. Figure 1 shows the scaled mean absolute error averaged over the four synthetic datasets. Figure 2 shows the results for the group error ratio.

Analyzing the results of the mean estimation, we see that all settings of k perform similarly well, especially when considering the large standard deviation of the different variants. However, there is a slight trend towards a better performance for larger k for larger ε and a worse performance for larger k for smaller ε. This is most obvious for the dataset with 2 groups and becomes less pronounced for the datasets with 8 and 64 groups, where the difference in performance between the different settings of k is almost negligible. The overall magnitude of the error is very similar for the different number of groups, but the standard deviation increases as the number of groups increases.

The discretization parameter k not only influences the error of the mean estimation, but also the amount of errors in the group information. The group error therefore differs for the different settings of k. While all parameter settings show a similarly large group error for small ε, the group errors deviate more for larger ε. While all settings of k show a decreasing group error for increasing ε, the group error is generally smaller for smaller k.

5.2.2 Choosing ε_1 and ε_2 for the Piecewise Mechanism. The Piecewise mechanism has a total privacy budget of $\varepsilon = \varepsilon_1 + \varepsilon_2$ and therefore requires a split of ε into ε_1 and ε_2. We investigate the influence of this split on the mean estimation error. Tables 3, 4, and 5 show the mean scaled absolute error of the Piecewise mechanism for different settings of ε_1, ε and $\varepsilon_2 = \varepsilon - \varepsilon_1$ for 2, 8, and 64 groups, respectively. For $\varepsilon > 1$, the best performing ratio $\varepsilon_1/\varepsilon$ is between 0.3 and 0.4 for 2 groups, while for 8 groups it is between 0.4 and 0.5, and for 64 groups it is between 0.6 and 0.7. For $\varepsilon \leq 1$, the best performing ratio is more erratic and does not follow the same trend as for larger ε. The error in this regime is generally very large and does not produce a meaningful result. Considering the trend for larger ε, we can conclude that the optimal split of ε into ε_1 and ε_2 depends on the number of groups and a larger number of groups requires a larger ε_1. However, in all cases, the best performing setting shows only a small improvement over the default choice of $\varepsilon_1 = \varepsilon_2 = \varepsilon/2$, especially when considering the large standard deviation of the errors (often on the order of magnitude of the error itself; see Tables 8, 9, and 10 in the appendix). Setting $\varepsilon_1 = \varepsilon_2 = \varepsilon/2$ is therefore a meaningful default, as it leads to a similar performance as the best performing setting for the different group sizes and does not require any additional tuning.

5.2.3 Estimating the Mean. In this experiment, we evaluate the accuracy of the four group mean estimation methods (Group Bernoulli, Group Laplace, Group Piecewise, and Group NPRR) on the synthetic datasets. We use fixed group sizes of 10 000 participants and vary the input range ($[-1, 1]$, $[-50, 50]$, $[0, 100]$) and the number of groups ($\{2, 8, 64\}$). Each simulation run was repeated 200 times to account for the randomness in the mechanism. Figure 3 shows the results averaged over the four synthetic datasets. Figure 4 shows the results for the group error ratio. We only show the results aggregated over all input ranges, as the different input ranges did not lead to a noticeable difference in the error.

Similar to the previous experiment, the standard deviations of the errors are quite large and overlap for all methods. Therefore, there is no clear best-performing method. Instead, we observe the following

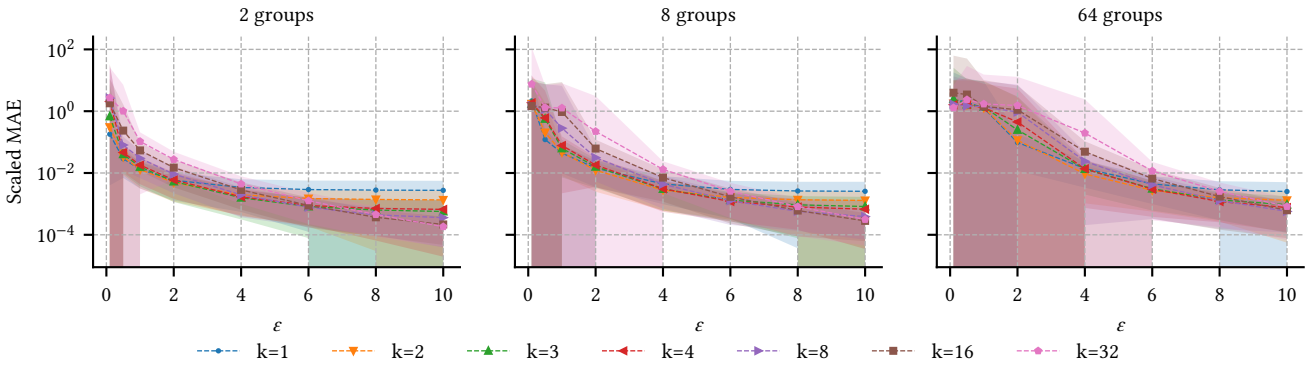


Figure 1: Scaled absolute error (absolute error divided by the data range of the dataset) for the Group NPRR method with different settings of k. The group size in all three subfigures is equal to 10 000. Shaded regions indicate the standard deviation over the randomness of the mechanism. The data for this figure was averaged over all four synthetic datasets (see Figure 7 in the appendix for detailed figures). The Figure shows that the error is generally smaller for larger k at larger ε and larger for smaller k at smaller ε, but the difference is small.

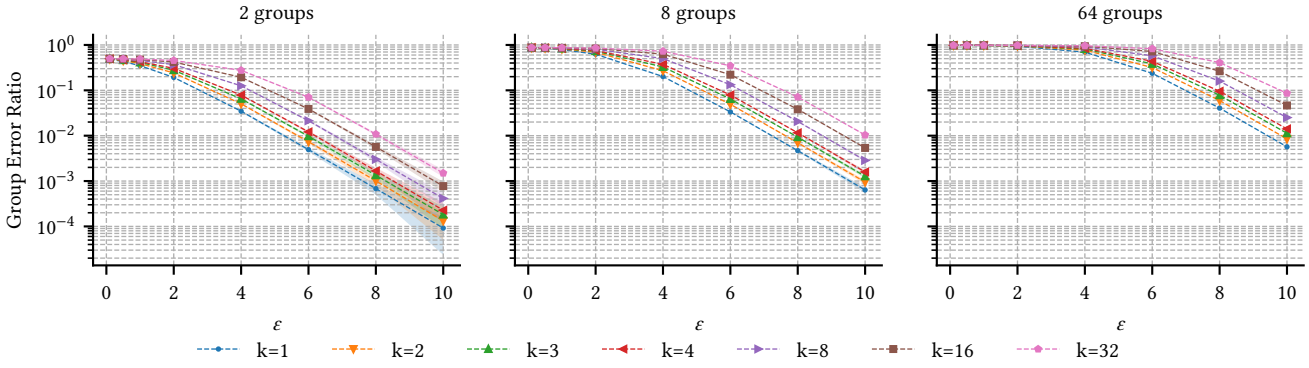


Figure 2: Group error ratio for the Group NPRR method with different settings of k. The group error ratio gives the number of incorrect group responses divided by the number of participants. The experiment uses the same setting as Figure 1. The Figure shows that the group error ratio is smaller for smaller k and shrinks with increasing ε. An increased number of groups leads to a higher group error.

Table 3: Mean scaled absolute error (absolute error divided by the data range of the dataset) of the Piecewise mechanism for different settings of ε_1, ε and $\varepsilon_2 = \varepsilon - \varepsilon_1$ for 2 groups averaged over the four synthetic datasets. Each group consists of 10 000 participants. The lowest error for each ε is shown in bold.

$\varepsilon_1/\varepsilon$ ε	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	7.51×10^{1}	1.55×10^{1}	1.71	$\mathbf{3.91\times10^{-1}}$	4.06×10^{-1}	4.67×10^{-1}	6.13×10^{-1}	9.29×10^{-1}	1.80
0.5	5.71×10^{-2}	$\mathbf{4.56\times10^{-2}}$	4.90×10^{-2}	5.38×10^{-2}	6.39×10^{-2}	7.80×10^{-2}	1.04×10^{-1}	1.57×10^{-1}	3.01×10^{-1}
1.0	2.38×10^{-2}	$\mathbf{2.04\times10^{-2}}$	2.18×10^{-2}	2.39×10^{-2}	2.74×10^{-2}	3.33×10^{-2}	4.34×10^{-2}	6.41×10^{-2}	1.28×10^{-1}
2.0	1.04×10^{-2}	$\mathbf{8.68\times10^{-3}}$	8.79×10^{-3}	9.53×10^{-3}	1.13×10^{-2}	1.37×10^{-2}	1.79×10^{-2}	2.58×10^{-2}	5.20×10^{-2}
4.0	4.68×10^{-3}	3.43×10^{-3}	$\mathbf{3.37\times10^{-3}}$	3.58×10^{-3}	4.28×10^{-3}	5.40×10^{-3}	7.04×10^{-3}	1.09×10^{-2}	2.28×10^{-2}
6.0	2.97×10^{-3}	2.03×10^{-3}	$\mathbf{1.92\times10^{-3}}$	2.03×10^{-3}	2.31×10^{-3}	3.06×10^{-3}	4.30×10^{-3}	6.72×10^{-3}	1.45×10^{-2}
8.0	2.22×10^{-3}	1.40×10^{-3}	$\mathbf{1.21\times10^{-3}}$	1.28×10^{-3}	1.58×10^{-3}	2.07×10^{-3}	2.98×10^{-3}	4.72×10^{-3}	1.05×10^{-2}
10.0	1.79×10^{-3}	1.03×10^{-3}	$\mathbf{8.21\times10^{-4}}$	8.72×10^{-4}	1.11×10^{-3}	1.55×10^{-3}	2.21×10^{-3}	3.69×10^{-3}	8.15×10^{-3}

Table 4: Mean scaled absolute error (absolute error divided by the data range of the dataset) of the Piecewise mechanism for different settings of ε_1, ε and $\varepsilon_2 = \varepsilon - \varepsilon_1$ for 8 groups averaged over the four synthetic datasets. Each group consists of 10 000 participants. The lowest error for each ε is shown in bold.

$\varepsilon_1/\varepsilon$ ε	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	$\mathbf{6.66\times10^{-1}}$	1.42	4.13	3.43	4.48	5.86	8.49	3.48×10^{1}	2.33×10^{1}
0.5	7.01×10^{-1}	8.89×10^{-1}	5.11×10^{-1}	3.17×10^{-1}	$\mathbf{2.61\times10^{-1}}$	2.91×10^{-1}	3.70×10^{-1}	5.33×10^{-1}	1.01
1.0	5.54×10^{-1}	1.30×10^{-1}	8.72×10^{-2}	$\mathbf{8.56\times10^{-2}}$	9.28×10^{-2}	1.07×10^{-1}	1.33×10^{-1}	1.85×10^{-1}	3.55×10^{-1}
2.0	1.14×10^{-1}	3.72×10^{-2}	3.02×10^{-2}	$\mathbf{2.88\times10^{-2}}$	2.97×10^{-2}	3.22×10^{-2}	3.87×10^{-2}	5.24×10^{-2}	9.67×10^{-2}
4.0	2.57×10^{-2}	1.30×10^{-2}	9.49×10^{-3}	8.03×10^{-3}	$\mathbf{7.55\times10^{-3}}$	8.07×10^{-3}	9.72×10^{-3}	1.36×10^{-2}	2.62×10^{-2}
6.0	1.51×10^{-2}	7.04×10^{-3}	4.54×10^{-3}	3.63×10^{-3}	$\mathbf{3.38\times10^{-3}}$	3.73×10^{-3}	4.77×10^{-3}	7.23×10^{-3}	1.49×10^{-2}
8.0	1.03×10^{-2}	4.44×10^{-3}	2.72×10^{-3}	2.05×10^{-3}	$\mathbf{1.94\times10^{-3}}$	2.29×10^{-3}	3.17×10^{-3}	5.02×10^{-3}	1.06×10^{-2}
10.0	7.80×10^{-3}	3.17×10^{-3}	1.80×10^{-3}	1.29×10^{-3}	$\mathbf{1.28\times10^{-3}}$	1.63×10^{-3}	2.33×10^{-3}	3.84×10^{-3}	8.39×10^{-3}

Table 5: Mean scaled absolute error (absolute error divided by the data range of the dataset) of the Piecewise mechanism for different settings of ε_1, ε and $\varepsilon_2 = \varepsilon - \varepsilon_1$ for 64 groups averaged over the four synthetic datasets. Each group consists of 10 000 participants. The lowest error for each ε is shown in bold.

$\varepsilon_1/\varepsilon$ ε	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	$\mathbf{6.21\times10^{-1}}$	1.59	2.37	3.20	7.60×10^{1}	7.37	1.15×10^{1}	4.28×10^{2}	4.35×10^{1}
0.5	8.62	**1.18**	2.08	2.64	4.54	5.50	1.70×10^{1}	1.41×10^{1}	1.17×10^{2}
1.0	$\mathbf{5.64\times10^{-1}}$	1.04	1.54	2.17	2.40	3.25	4.42	4.08	1.53×10^{1}
2.0	5.48×10^{-1}	9.50×10^{-1}	1.07	8.49×10^{-1}	3.31×10^{-1}	$\mathbf{2.37\times10^{-1}}$	2.44×10^{-1}	3.05×10^{-1}	5.17×10^{-1}
4.0	6.98×10^{-1}	5.49×10^{-1}	8.13×10^{-2}	4.86×10^{-2}	3.78×10^{-2}	3.27×10^{-2}	$\mathbf{3.20\times10^{-2}}$	3.69×10^{-2}	5.94×10^{-2}
6.0	7.50×10^{-1}	6.54×10^{-2}	2.72×10^{-2}	1.64×10^{-2}	1.12×10^{-2}	8.89×10^{-3}	$\mathbf{8.55\times10^{-3}}$	1.03×10^{-2}	1.86×10^{-2}
8.0	4.76×10^{-1}	3.05×10^{-2}	1.35×10^{-2}	7.26×10^{-3}	4.65×10^{-3}	$\mathbf{3.71\times10^{-3}}$	3.93×10^{-3}	5.50×10^{-3}	1.12×10^{-2}
10.0	1.52×10^{-1}	1.87×10^{-2}	7.59×10^{-3}	3.83×10^{-3}	2.42×10^{-3}	$\mathbf{2.08\times10^{-3}}$	2.53×10^{-3}	3.92×10^{-3}	8.46×10^{-3}

trends: The Group NPRR method performs best for large ε (see also the previous experiment for the setting of k), whereas the Group Bernoulli method (equivalent to NPPR with $k = 1$) performs best for small ε, closely followed by the Group Laplace and Group Piecewise methods. The performance of these three methods levels off quickly for larger ε, with the Group Bernoulli method showing the largest error for large ε, which resembles their behavior in the non-group setting (see Raab et al. [39] for detailed results in the non-group setting). Examining the detailed results for the various datasets (refer to Figure 9 in the appendix), we observe similar trends across all datasets, except for the Extremum dataset. In the Extremum dataset, the Group Bernoulli method outperforms the other methods for all ε settings. This behavior can be attributed to the fact that the Extremum dataset contains only the two extreme values s and r, which align with the values used by the Group Bernoulli method for rounding. Consequently, the rounding procedure does not introduce any additional error in this case.

Examining the group error, we observe that all methods exhibit a similar group error ratio for $\varepsilon = 0.1$, which decreases as ε increases. The Group Bernoulli method and the NPRR variants follow the same patterns as in the previous experiment, with Bernoulli (NPRR with $k = 1$) showing the lowest group error and NPRR with $k = 32$ showing the highest. Both the Piecewise and Laplace methods, which use $\varepsilon_1 = \varepsilon/2$, display the same group error that gradually decreases with larger ε. Generally, the group error is higher for a greater number of groups, and the reduction in group error becomes noticeable only at larger ε values.

5.2.4 Impact of Group Imbalance. To examine the effect of group imbalance, we use synthetic datasets with two groups of varying sizes. We simulate datasets where the first group has $n_0 = 10\,000$ participants, and the size ratios between the first and second group are $n_0/n_1 = \{1, 0.1, 0.01\}$. This means the second group has sizes of $\{10\,000, 100\,000, 1\,000\,000\}$ participants.

Figure 5 presents the results for mean estimation across different methods. The error patterns for the various methods are similar to those observed in balanced groups in the previous section. The error magnitude does not change substantially with different imbalance ratios, but the standard deviation is generally larger for imbalanced groups.

To explore the increased standard deviation further, we display the scaled absolute error for individual groups in Figure 11 in the appendix. This figure shows that the error for the larger group is generally smaller than that for the smaller group and that the error for the larger group decreases with increasing imbalance, while the error for the smaller group increases slightly. This indicates that the mean estimation error of one group is influenced not only by the group size but also by the size of the other group. However, this effect is rather small.

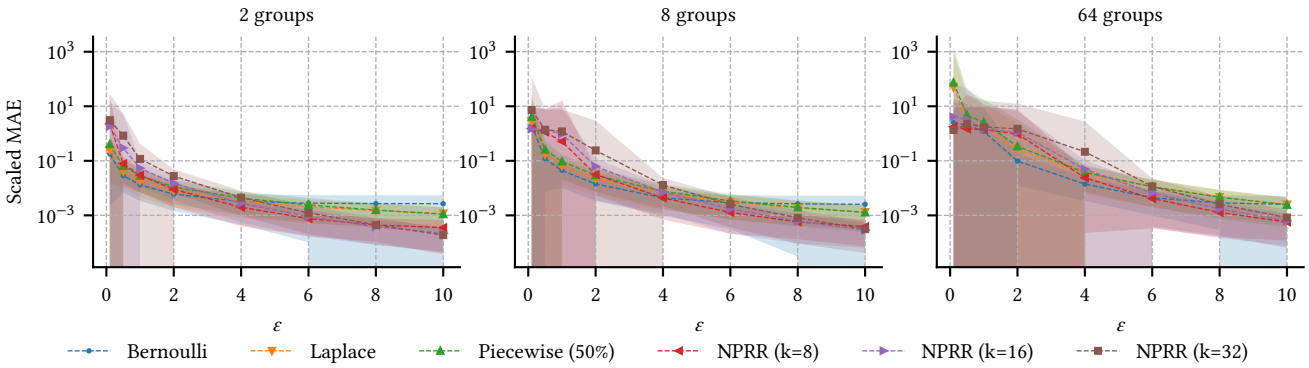


Figure 3: Scaled absolute error (absolute error divided by the data range of the dataset) for the different proposed methods. The group size in all three subfigures is equal to 10 000. Shaded regions indicate the standard deviation over the randomness of the mechanism. The data for this figure was averaged over all four synthetic datasets (see Figure 9 in the appendix for detailed figures). The Figure shows that most methods perform similarly well, with the Group NPRR method showing the best performance for large ε and the Group Bernoulli method showing the best performance for small ε.

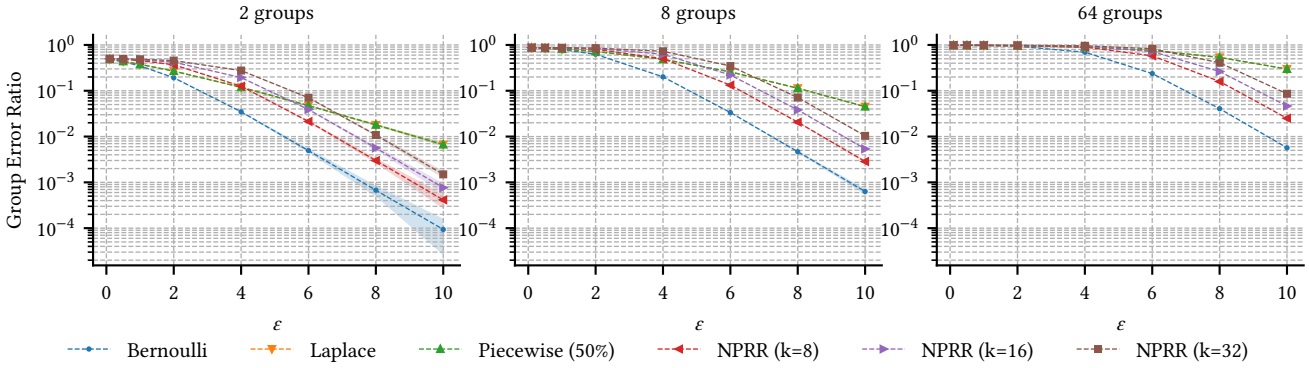


Figure 4: Group error ratio for the different proposed methods. The group error ratio gives the number of incorrect group responses divided by the number of participants. The experiment uses the same setting as Figure 3. The Figure shows that the group error ratio shrinks with increasing ε, but at different rates for the different methods.

5.3 Validation on Medical Data

In this section, we examine individual subsets of the AQUILA and MIMIC-IV datasets to assess the performance of the different mean estimation methods in a real-world scenario. To visualize the randomness in the mechanisms, we simulate each setting 200 times. For each dataset, we plot the individual mean estimation results and the true mean for each group across various ε settings ($\{0.5, 1.0, 4.0, 10.0\}$). We focus on the variance of the estimated means and their proximity to the true mean, as these factors are crucial for the utility of the methods and their potential application in medical data analysis. If the variance is too high or the estimated mean deviates from the true mean, the results must be considered unreliable. The magnitude of the variance or deviation that can be considered acceptable in practice depends on the concrete use case and further down-stream analysis questions. For this reason, we keep this analysis rather vague and present the results and their relative comparison.

5.3.1 AQUILA. For the AQUILA dataset, we discuss the two combinations *mean age per disease* and *mean height per gender* in more detail and show further results in the appendix (Figures 12 and 13). Figure 6a shows the results for the *mean age per disease* and Figure 6b shows the results for the *mean height per gender*.

For groups defined by the disease attribute, there are roughly twice as many participants with PsA compared to AS, while the groups are almost balanced for the gender attribute. Despite this imbalance, both figures exhibit similar behavior across different ε settings. For small ε values (0.5 or 1.0), the estimated means show high variance and often deviate significantly from the true mean, sometimes spanning the entire input range for $\varepsilon = 0.5$. As ε increases (2.0 or 4.0), the variance decreases, and the estimated means become closer to the true mean. The different methods display similar trends as observed with synthetic data: the Bernoulli method has the smallest variance for small ε but the largest for large ε, while the Laplace and Piecewise methods maintain consistent variance with a slight decrease for larger ε. The NPRR method shows very high

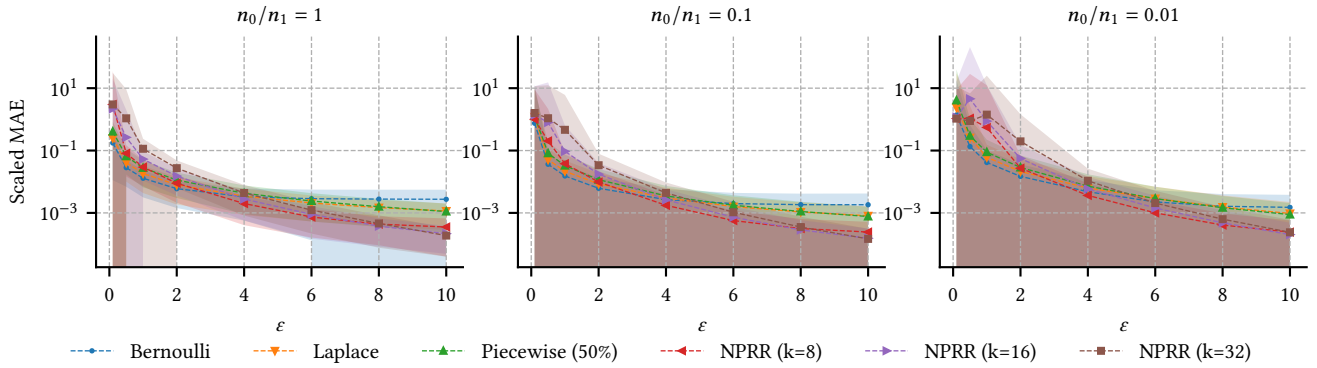


Figure 5: Scaled absolute error (absolute error divided by the data range of the dataset) for the different proposed methods for imbalanced groups. The group size for the first group is equal to 10 000. The second group has 10 000 participants in the first subfigure, 100 000 participants in the second subfigure, and 1 000 000 participants in the third subfigure. Shaded regions indicate the standard deviation over the randomness of the mechanism. The figure shows that there is no substantial difference in the error magnitude for different imbalance ratios, but the standard deviation is generally larger for more imbalanced groups.

variance for small ε but the smallest variance for larger ε values (4.0 and 10.0).

Next to the actual mean estimates, we are also interested in the difference between the means of two groups. We show the results for the difference between the means of the two groups for the AQUILA dataset in Figures 15 and 16 in the appendix. The violin plots illustrate the distribution of the random results for the difference between the means of the two groups across various ε settings. In all cases, the means of the random results are close to the true difference between the means of the two groups, indicating that the mean estimation methods are unbiased. Furthermore, we observe similar trends as in the previous plots, with the NPRR method exhibiting a very large variance for small ε and a smaller variance compared to the other methods for larger ε. Additionally, Figures 18 and 19 illustrate the frequency with which the order of means was correctly estimated across 200 random runs. These plots indicate that the accuracy of mean order estimation improves with increasing ε, with most methods achieving a 100% accuracy for datasets where the difference between group means is substantial. For smaller differences, the Bernoulli method occasionally fails to estimate the correct order of means, not reaching a 100% accuracy even for $\varepsilon = 10.0$. For all methods, the ε value of first reaching a 100% accuracy varies depending on the difference between the group means. In the case of the *BMI per gender* dataset, the group difference is so small, that no method consistently estimates the correct order for all random runs, even at the highest ε settings.

5.3.2 MIMIC-IV. For the MIMIC-IV dataset, we discuss the two combinations *mean age per gender* and *mean age per admission location* in more detail and show further results in the appendix (Figure 14). Figure 6c shows the results for the *mean age per gender* and Figure 6d shows the results for the *mean age per admission location*.

For the groups defined by the gender attribute, we have roughly balanced groups (with roughly 10% more female than male). However, for the admission location groups, there is a large imbalance with most participants (232 476) being in group 2, a few thousand each in group 1, 4, 5, 7, and 9, and only a few hundred in groups 0 and 3.

For the case where groups are defined by the gender, we see a very similar behavior to the AQUILA dataset, but with a much smaller variance in the estimated means.

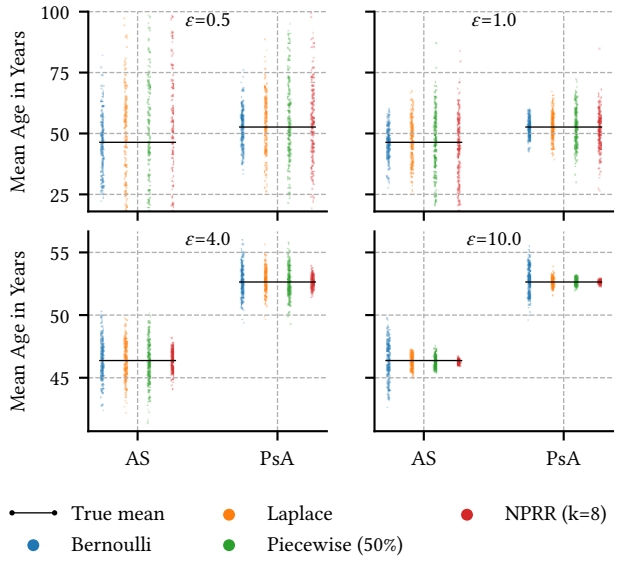
For the admission location, we see a large variance in the estimated means for small ε (0.5 and 1.0) and a smaller variance for larger ε (2.0 and 4.0). For $\varepsilon = 0.5$, only the two largest groups show a variance in the mean estimate that is not covering the full input range. For $\varepsilon = 1.0$, the variance decreases and the estimated means are closer to the true mean. For $\varepsilon = 4.0$ and $\varepsilon = 10.0$, the variance for most groups is very small and only the smallest groups still show a larger variance.

The fact that the variance is still large for $\varepsilon = 1.0$ points towards the fact that not only the number of participants in the group but also the number of groups has an impact on the variance of the estimated means.

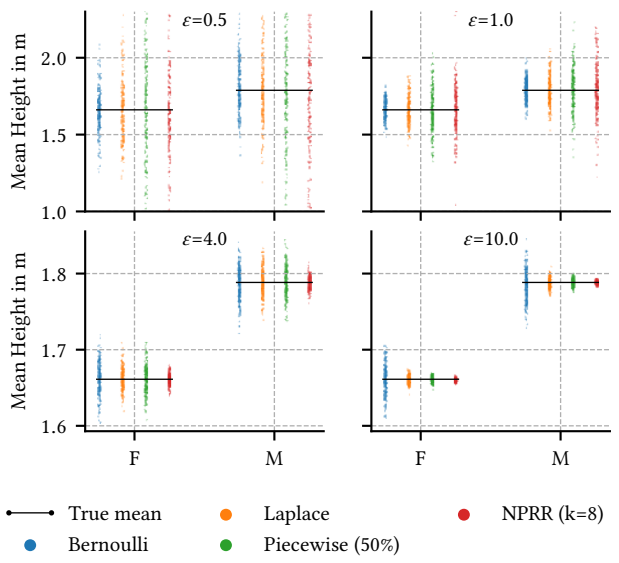
We show the results for the difference between the means of the two groups for the MIMIC-IV dataset in Figures 17 in the appendix. The results are similar to the AQUILA dataset, with the means of the simulation results being close to the true difference between the means of the two groups. This indicates that the mean estimation methods are unbiased. As the number of participants in the groups is much larger for the MIMIC-IV dataset, the variance of the estimated means is much smaller compared to the AQUILA dataset. We also see this in the ratio of the correct order of the means in Figures 20 in the appendix, where the order of the means is estimated correctly in almost all cases. The order ratio is only smaller than 100% for the smallest ε settings.

6 Discussion

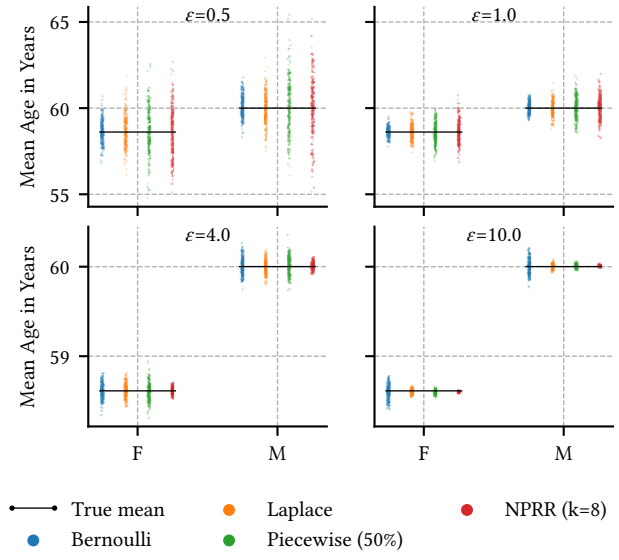
In this paper, we have introduced and evaluated several methods for estimating the mean of a group of participants in a decentralized setting where both the group and the value are considered private.



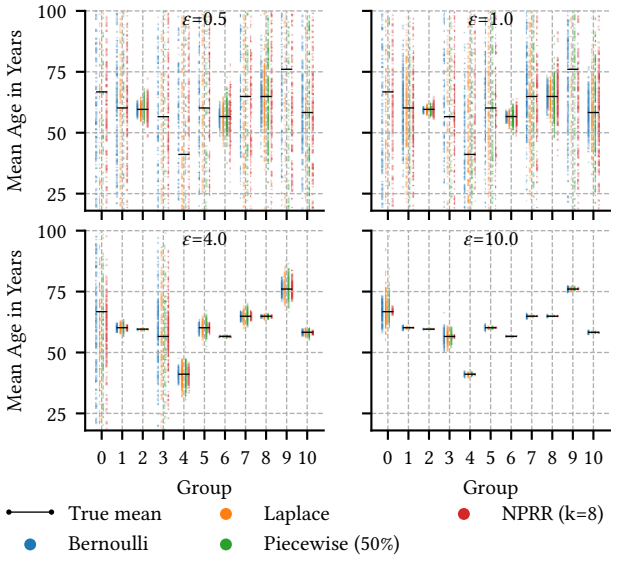
(a) AQUILA dataset. Group: Disease, Value: Age.
Participants per group: AS: 683, **PsA:** 1278



(b) AQUILA dataset. Group: Gender, Value: Height.
Participants per group: F: 993, **M:** 921



(c) MIMIC-IV dataset. Group: Gender, Value: Age.
Participants per group: F: 224 796, **M:** 206 066



(d) MIMIC-IV dataset. Group: Admission Location, Value: Age.
Participants per group: 0: 185, **1:** 10 006, **2:** 232 476, **3:** 358, **4:** 4204, **5:** 5401, **6:** 114 815, **7:** 7798, **8:** 35 964, **9:** 3842, **10:** 15 813
See Table 7 in the appendix for group descriptions.

Figure 6: Individual results for the mean estimation for different groups and values using data from the AQUILA and MIMIC-IV datasets. Each method was executed 200 times and each resulting mean estimate is shown as a single point to visualize the randomness of the mechanism. The true mean is shown as a solid line. All subfigures show that the variance of the estimated means generally decreases with increasing ε. The different methods exhibit similar trends, with the NPRR method performing best for larger ε and the Bernoulli method performing best for smaller ε. The variance of the estimated means is generally higher for smaller groups and smaller ε.

6.1 General Findings

We have seen that the accuracy of the estimated means depends largely on the number of participants in the corresponding group, the total number of groups and the value of ε. It also depends on the number of participants in the other groups, but to a lesser extent. When groups have many participants and ε is large, group means can be estimated quite accurately, but when they are not large enough or there are too many groups, the estimated means resemble a uniform sampling from the relevant input domain and are therefore not very useful. These results are generally in line with the findings for mean estimation methods which do not consider group information [39]. One notable exception is the Piecewise mechanism, which does not perform as well as in the non-group setting [39].

The results of the experiments on the MIMIC-IV dataset show that the methods can be applied to real-world medical data and provide meaningful results, given enough participants in each group combined with a reasonably large value of ε. For the smaller AQUILA dataset, the results are less reliable, as the number of participants per group is smaller and a larger value of ε is required to obtain accurate estimates. Given the requirement of a large value of ε for accurate estimates, a method should be chosen that works well with a large value of ε such as NPRR with a large k.

6.2 Limitations

A common challenge with LDP methods is the requirement for a large dataset to achieve accurate estimates, which we also observed in our experiments. As discussed above, the accuracy of the estimated means depends on the number of participants in the corresponding group, the total number of groups, and the value of ε. When the number of participants in the groups is small or the number of groups is large, the accuracy of the mean estimates deteriorates. This limitation is inherent to the LDP framework and is not specific to the methods proposed in this paper.

Specifically, the group-based mean estimation methods proposed in this paper are particularly affected when the number of groups is large or the number of participants per group is small. In such cases, the GRR method used for estimating the group information may not yield accurate results, which in turn affects the accuracy of the mean estimates. Considering the error introduced by the group information, alternative methods from frequency estimation, such as k-Subset [43, 48] or RAPPOR [20], could potentially yield better results compared to GRR. However, these alternatives require a thorough investigation of their privacy guarantees and the resulting mean estimators, which we leave for future work.

Additionally, the Piecewise mechanism specifically presents a problem in deriving an optimal setting for the privacy parameters ε_1 and ε_2. As discussed in Section 5.2.2, setting these parameters to $\varepsilon/2$ is an acceptable default, but not optimal. Further research is needed to develop strategies for optimally setting these parameters in different scenarios or to find a theoretically optimal ratio.

6.3 Towards Application

Before the methods in this paper can be used in a real-world application, several challenges need to be addressed by practitioners.

6.3.1 Choosing the Input Range. All discussed group mean estimation methods, like most other LDP methods, require a bounded input domain to function well. In our experiments, we predetermined the size of the input range. This presents two challenges: If the input range is too small, many data points become "outliers" and must be projected to the boundaries of the input range, negatively impacting the quality of the mean estimate. Conversely, if the input range is too large, the accuracy of the mean estimate deteriorates, as the error magnitude increases linearly with the input range size – similar to the non-group setting [39].

We propose a potential solution to this issue and leave the detailed analysis for future work. One possible approach is to implement a binary search for the input domain. The idea is to start with a very large input range and then iteratively narrow it down by asking participants whether their data falls within the current range. Participants use randomized response to perturb their answers, and frequency estimation is used to estimate the proportion of participants within the chosen range. This process continues until the estimated proportion of participants within the range falls below a certain threshold.

According to the composition theorem, summing the ε values of individual queries in pure LDP results in a large overall ε, weakening the privacy guarantee. A potential solution could be the use of the shuffle model of differential privacy [7]. In this model, the shuffler collects the data from the participants and shuffles it before sending it to the analyst. By removing the connection between the participants and their responses, the shuffle model can provide a stronger privacy guarantee than pure LDP. Furthermore, the shuffle model allows each query to use the full ε without additional privacy loss [19]. This raises the question whether this method is only applicable in the shuffle model, or if it can also be adapted for the local model. Further research is needed to explore the feasibility and effectiveness of this approach in both models, potentially leading to more robust and practical privacy-preserving data analysis techniques.

6.3.2 Large Privacy Budgets. As we have seen in the experimental results, the privacy budget ε required for accurate mean estimates can be quite large. In the pure LDP setting, this is not desirable, as the privacy guarantee becomes weaker (or even meaningless) with larger values of ε. Furthermore, following the composition theorem, the privacy budget needs to be summed up over multiple queries, which can quickly lead to a very large value of ε for which the LDP guarantee is not very strong (or even meaningless).

Possible solutions to this problem include the use of other notions of privacy (see next section) or the use of the shuffle model of differential privacy. This would allow for a larger value of ε for the LDP mechanism while still achieving a good DP protection for the resulting means.

Furthermore, the shuffle model also allows for multiple queries for different attributes without having to sum up the privacy budget [19], which is a significant advantage over the pure LDP setting. Since the shuffle model directly builds on the LDP model, the methods introduced in this paper can be directly applied in the shuffle model without any modifications.

To actually provide additional trust, the shuffler needs to be operated by trusted entities, such as hospitals or doctors in the

healthcare domain, who are already trusted by patients to handle their data. We leave the application of the shuffle model and the evaluation of the methods for this model for future work.

6.3.3 Other Notions of Privacy. While this work focuses on pure LDP, other variants exist. Metric LDP, for example, relaxes the indistinguishability requirement over the entire domain, allowing for more accurate data analysis while still protecting privacy [3]. This approach can be applied in any domain with a defined metric and offers a better privacy-utility trade-off, especially with advanced statistical utility measures like the earth mover's distance [2, 6, 18]. However, applying metric LDP to group mean estimation, particularly for medical data, requires further research into adaptations of methods, the correct selection of metrics, and a careful analysis of the privacy guarantees and utility of the resulting estimators.

6.3.4 Need for Further Analysis Methods for Medical Data. While calculating the means of groups is a common task in medical data analysis, it is not sufficient to fully understand the underlying data. We discuss some additional methods in the following.

Frequency Estimation with Groups. Similar to the methods proposed in this paper, it is possible to estimate value frequencies over multiple private groups. This can be done by replacing the mean estimation component of our generic algorithm (Algorithm 3) with one of the many existing frequency estimation methods. However, the problem of estimating frequencies for multiple groups is also equivalent to estimating the joint distribution between the categorical groups and values, which has also been discussed in LDP literature [8, 22, 22, 47, 49]. Indeed, the joint distribution estimation algorithm by Xue et al. [47] outperformed a combination of GRR for both groups and values in preliminary experiments. We leave a detailed comparison for future work.

Comparing Group Means with Significance. In medical data analysis, it is common to use statistical tests to determine if differences in group means are significant. Ding et al. [11] have shown how to perform hypothesis tests to compare the mean between two groups by using their Bernoulli mechanism [10]. However, as stated in the related work, this method does not consider the group as private. Additionally, Waudby-Smith et al. [46] show how to estimate the mean and confidence intervals for the population mean on the basis of their NPRR method. Building on the group-wise mean estimation and the ideas of these works, developing methods for comparing the means of different groups with significance in a privacy-preserving manner is a promising direction for future research.

7 Conclusion

Given the developments in the European Union towards the large-scale secondary use of health data for research and development, there is an urgent need for proper privacy-preserving data analysis procedures. In this paper, we discussed the estimation of group means as a central data analysis task and discussed four LDP methods for this task. We showed their privacy guarantees, and evaluated their performance on synthetic and real-world medical datasets. While these methods provide a foundation for data analysis in the EHDS, further methods are needed to make LDP applicable to practical medical data analysis.

Going forward, differential privacy should certainly play a role in the context of the EHDS (to protect published aggregation results from attacks). However, LDP may not be feasibly applied in all situations (especially when the amount of available data is small or high accuracy is required). In these cases, combinations with the shuffle model of differential privacy or cryptographic solutions like secure multi-party-computation should be considered. Relaxations of the privacy guarantee (e.g., (ε, δ)-(L)DP, or metric LDP) could also be considered for practical applications. However, the impact of these relaxations on the privacy guarantee and the accuracy of the results should be carefully evaluated.

Acknowledgments

We thank the anonymous reviewers and the revision editor for their helpful comments and suggestions. René Raab acknowledges funding provided by the Federal Ministry for Economic Affairs and Climate Action (BMWK) under Grant No. 68GX21004F (TEAM-X).

We thank Novartis for providing access to the AQUILA study data used in this work. The authors gratefully acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). The hardware is funded by the German Research Foundation (DFG).

References

[1] Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. 2019. Hadamard Response: Estimating Distributions Privately, Efficiently, and with Little Communication. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. PMLR, 1120–1129. https://proceedings.mlr.press/v89/acharya19a.html

[2] Mário Alvim, Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Anna Pazii. 2018. Invited Paper: Local Differential Privacy on Metric Spaces: Optimizing the Trade-Off with Utility. In *2018 IEEE 31st Computer Security Foundations Symposium (CSF)*. 262–267. https://doi.org/10.1109/CSF.2018.00026

[3] Mário S. Alvim, Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Anna Pazii. 2018. Metric-Based Local Differential Privacy for Statistical Applications. https://doi.org/10.48550/arXiv.1805.01456 arXiv:1805.01456 [cs]

[4] Raef Bassily and Adam Smith. 2015. Local, Private, Efficient Protocols for Succinct Histograms. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing (STOC '15)*. Association for Computing Machinery, New York, NY, USA, 127–135. https://doi.org/10.1145/2746539.2746632

[5] Pascal Berrang, Paul Gerhart, and Dominique Schröder. 2024. Measuring Conditional Anonymity — A Global Study. *Proceedings on Privacy Enhancing Technologies* 2024, 4 (2024), 947–966. https://doi.org/10.56553/popets-2024-0150

[6] Sayan Biswas and Catuscia Palamidessi. 2024. PRIVIC: A Privacy-Preserving Method for Incremental Collection of Location Data. *Proceedings on Privacy Enhancing Technologies* (2024). https://doi.org/10.56553/popets-2024-0033

[7] Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Usharsee Kode, Julien Tinnes, and Bernhard Seefeld. 2017. Prochlo: Strong Privacy for Analytics in the Crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles*. 441–459. https://doi.org/10.1145/3132747.3132769 arXiv:1710.00901 [cs]

[8] Graham Cormode, Tejas Kulkarni, and Divesh Srivastava. 2018. Marginal Release Under Local Differential Privacy. In *Proceedings of the 2018 International Conference on Management of Data*. ACM, Houston TX USA, 131–146. https://doi.org/10.1145/3183713.3196906

[9] Graham Cormode, Samuel Maddock, and Carsten Maple. 2021. Frequency Estimation under Local Differential Privacy. *Proceedings of the VLDB Endowment* 14, 11 (July 2021), 2046–2058. https://doi.org/10.14778/3476249.3476261

[10] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. 2017. Collecting Telemetry Data Privately. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/hash/253614bbac999b38b5b60cae531c4969-Abstract.html

[11] Bolin Ding, Harsha Nori, Paul Li, and Joshua Allen. 2018. Comparing Population Means Under Local Differential Privacy: With Significance and Power. *Proceedings of the AAAI Conference on Artificial Intelligence* 32, 1 (April 2018). https://doi.org/10.1609/aaai.v32i1.11301

[12] John Duchi, Martin J Wainwright, and Michael I Jordan. 2013. Local Privacy and Minimax Bounds: Sharp Rates for Probability Estimation. In *Advances in Neural Information Processing Systems*, Vol. 26. Curran Associates, Inc. https://papers.nips.cc/paper_files/paper/2013/hash/5807a685d1a9ab3b599035bc566ce2b9-Abstract.html

[13] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. 2014. Local Privacy, Data Processing Inequalities, and Statistical Minimax Rates. arXiv:1302.3203 [cs, math, stat] http://arxiv.org/abs/1302.3203 arXiv:1302.3203.

[14] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. 2018. Minimax Optimal Procedures for Locally Private Estimation. *J. Amer. Statist. Assoc.* 113, 521 (Jan. 2018), 182–201. https://doi.org/10.1080/01621459.2017.1389735

[15] Cynthia Dwork and Jing Lei. 2009. Differential Privacy and Robust Statistics. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing (STOC '09)*. Association for Computing Machinery, New York, NY, USA, 371–380. https://doi.org/10.1145/1536414.1536466

[16] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *Theory of Cryptography (Lecture Notes in Computer Science)*, Shai Halevi and Tal Rabin (Eds.). Springer, Berlin, Heidelberg, 265–284. https://doi.org/10.1007/11681878_14

[17] Cynthia Dwork and Aaron Roth. 2014. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (Aug. 2014), 211–407. https://doi.org/10/gcgmcw

[18] Ehab ElSalamouny and Catuscia Palamidessi. 2020. Generalized Iterative Bayesian Update and Applications to Mechanisms for Privacy Protection. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*. 490–507. https://doi.org/10.1109/EuroSP48549.2020.00038

[19] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Shuang Song, Kunal Talwar, and Abhradeep Thakurta. 2020. Encode, Shuffle, Analyze Privacy Revisited: Formalizations and Empirical Evaluation. https://doi.org/10.48550/arXiv.2001.03618 arXiv:2001.03618 [cs]

[20] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14)*. Association for Computing Machinery, New York, NY, USA, 1054–1067. https://doi.org/10.1145/2660267.2660348

[21] European Commission, Directorate-General for Health and Food Safety. 2022. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space COM(2022) 197 Final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0197

[22] Giulia Fanti, Vasyl Pihur, and Úlfar Erlingsson. 2016. Building a RAPPOR with the Unknown: Privacy-Preserving Learning of Associations and Data Dictionaries. *Proceedings on Privacy Enhancing Technologies* 2016, 3 (July 2016), 41–61. https://doi.org/10.1515/popets-2016-0015

[23] Rina Friedberg and Ryan Rogers. 2023. Privacy Aware Experimentation over Sensitive Groups: A General Chi Square Approach. In *Proceedings of the Workshop on Algorithmic Fairness through the Lens of Causality and Privacy*. PMLR, 23–66. https://proceedings.mlr.press/v214/friedberg23a.html

[24] Marco Gaboardi, Ryan Rogers, and Or Sheffet. 2019. Locally Private Mean Estimation: Z-Test and Tight Confidence Intervals. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. PMLR, 2545–2554. https://proceedings.mlr.press/v89/gaboardi19a.html

[25] Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. 2000. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101, 23 (June 2000). https://doi.org/10.1161/01.CIR.101.23.e215

[26] Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. 2023. MIMIC-IV. https://doi.org/10.13026/6MM1-EK67

[27] Alistair Johnson, Tom Pollard, a-chahin, Jim Blundell, Brian Gow, erinhong, Michael Schubert, Kien Dang, Nicolas Paris, JackieMe, shu98, Eric Carlson, Andrew Barros, Qinyu Zhao, etheleon, Lucas Bulgarelli, Sicheng Hao, alexmbennett2, Christian Porschen, Stefan Hegselmann, Peter Szolovits, Hans0124SG, Armando Fandango, C. V. Cosgriff, Juan L. Domínguez-Olmedo, Sravan R, Thomas M Ward, Yong Fan, Pedro Gemal Lanzieri, and Thijs van den Berg. 2024. MIT-LCP/Mimic-Code: V2.5.0. Zenodo. https://doi.org/10.5281/ZENODO.13374956

[28] Alistair E. W. Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J. Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, Li-wei H. Lehman, Leo A. Celi, and Roger G. Mark. 2023. MIMIC-IV, a Freely Accessible Electronic Health Record Dataset. *Scientific Data* 10, 1 (Jan. 2023), 1. https://doi.org/10.1038/s41597-022-01899-x

[29] Alistair E W Johnson, David J Stone, Leo A Celi, and Tom J Pollard. 2018. The MIMIC Code Repository: Enabling Reproducibility in Critical Care Research. *Journal of the American Medical Informatics Association* 25, 1 (Jan. 2018), 32–39. https://doi.org/10.1093/jamia/ocx084

[30] Matthew Joseph, Janardhan Kulkarni, Jieming Mao, and Steven Z. Wu. 2019. Locally Private Gaussian Estimation. In *Advances in Neural Information Processing Systems*, Vol. 32. Curran Associates, Inc. https://papers.nips.cc/paper_files/paper/2019/hash/a588a6199feff5ba48402883d9b72700-Abstract.html

[31] Marc Juarez and Aleksandra Korolova. 2023. "You Can't Fix What You Can't Measure": Privately Measuring Demographic Performance Disparities in Federated Learning. In *Proceedings of the Workshop on Algorithmic Fairness through the Lens of Causality and Privacy*. PMLR, 67–85. https://proceedings.mlr.press/v214/juarez23a.html

[32] Peter Kairouz, Keith Bonawitz, and Daniel Ramage. 2016. Discrete Distribution Estimation under Local Privacy. In *Proceedings of The 33rd International Conference on Machine Learning*. PMLR, 2436–2444. https://proceedings.mlr.press/v48/kairouz16.html

[33] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. 2014. Extremal Mechanisms for Local Differential Privacy. In *Advances in Neural Information Processing Systems*, Vol. 27. Curran Associates, Inc. https://papers.nips.cc/paper_files/paper/2014/hash/86df7dcfd896fcaf2674f757a2463eba-Abstract.html

[34] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. 2016. Extremal Mechanisms for Local Differential Privacy. *Journal of Machine Learning Research* 17, 17 (2016), 1–51. http://jmlr.org/papers/v17/15-135.html

[35] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2011. What Can We Learn Privately? *SIAM J. Comput.* 40, 3 (Jan. 2011), 793–826. https://doi.org/10.1137/090756090

[36] Uta Kiltz, Carolin Legeler, Monika Maier-Peuschel, Christian Mann, and Hans-Peter Tony. 2019. Baseline Characteristics of Patients with Ankylosing Spondylitis and Psoriatic Arthritis Treated with Secukinumab in the Real-World Setting: AQUILA, a Non-Interventional Study. *The Open Rheumatology Journal* 13 (2019), 53–60. https://doi.org/10.2174/1874312901913010053

[37] Takao Murakami, Hideitsu Hino, and Jun Sakuma. 2018. Toward Distribution Estimation under Local Differential Privacy with Small Samples. *Proceedings on Privacy Enhancing Technologies* (2018). https://doi.org/10.1515/popets-2018-0022

[38] Thông T. Nguyên, Xiaokui Xiao, Yin Yang, Siu Cheung Hui, Hyejin Shin, and Junbum Shin. 2016. Collecting and Analyzing Data from Smart Device Users with Local Differential Privacy. https://doi.org/10.48550/arXiv.1606.05053 arXiv:1606.05053 [cs] arXiv:1606.05053.

[39] René Raab, Pascal Berrang, Paul Gerhart, and Dominique Schröder. 2025. SoK: Descriptive Statistics Under Local Differential Privacy. *Proceedings on Privacy Enhancing Technologies* (2025). https://doi.org/10.56553/popets-2025-0008

[40] René Raab, Arne Küderle, Anastasiya Zakreuskaya, Ariel D. Stern, Jochen Klucken, Georgios Kaissis, Daniel Rueckert, Susanne Boll, Roland Eils, Harald Wagener, and Bjoern M. Eskofier. 2023. Federated Electronic Health Records for the European Health Data Space. *The Lancet Digital Health* 5, 11 (Nov. 2023), e840–e847. https://doi.org/10.1016/S2589-7500(23)00156-5

[41] Latanya Sweeney. 1997. Weaving Technology and Policy Together to Maintain Confidentiality. *The Journal of Law, Medicine & Ethics* 25, 2-3 (June 1997), 98–110. https://doi.org/10.1111/j.1748-720X.1997.tb01885.x

[42] Ning Wang, Xiaokui Xiao, Yin Yang, Jun Zhao, Siu Cheung Hui, Hyejin Shin, Junbum Shin, and Ge Yu. 2019. Collecting and Analyzing Multidimensional Data with Local Differential Privacy. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE Computer Society, 638–649. https://doi.org/10.1109/ICDE.2019.00063

[43] Shaowei Wang, Liusheng Huang, Pengzhan Wang, Yiwen Nie, Hongli Xu, Wei Yang, Xiang-Yang Li, and Chunming Qiao. 2016. Mutual Information Optimally Local Private Discrete Distribution Estimation. https://doi.org/10.48550/arXiv.1607.08025 arXiv:1607.08025 [cs, math] arXiv:1607.08025.

[44] Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. 2017. Locally Differentially Private Protocols for Frequency Estimation. In *26th USENIX Security Symposium (USENIX Security 17)*. 729–745. https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/wang-tianhao

[45] Stanley L. Warner. 1965. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *J. Amer. Statist. Assoc.* 60, 309 (March 1965), 63–69. https://doi.org/10.1080/01621459.1965.10480775

[46] Ian Waudby-Smith, Steven Wu, and Aaditya Ramdas. 2023. Nonparametric Extensions of Randomized Response for Private Confidence Sets. In *Proceedings of the 40th International Conference on Machine Learning*. PMLR, 36748–36789. https://proceedings.mlr.press/v202/waudby-smith23a.html

[47] Qiao Xue, Youwen Zhu, and Jian Wang. 2021. Joint Distribution Estimation and Naïve Bayes Classification Under Local Differential Privacy. *IEEE Transactions on Emerging Topics in Computing* 9, 4 (Oct. 2021), 2053–2063. https://doi.org/10.1109/TETC.2019.2959581

[48] Min Ye and Alexander Barg. 2017. Optimal Schemes for Discrete Distribution Estimation under Local Differential Privacy. In *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, Aachen, Germany, 759–763. https://doi.org/10.1109/ISIT.2017.8006630

[49] Zhikun Zhang, Tianhao Wang, Ninghui Li, Shibo He, and Jiming Chen. 2018. CALM: Consistent Adaptive Local Marginal for Marginal Release under Local Differential Privacy. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18)*. Association for Computing Machinery, New York, NY, USA, 212–229. https://doi.org/10.1145/3243734.3243742

A Additional Figures & Tables

Table 6: Notation used in this paper

Symbol	Description
$\mathbf{1}_A$	Indicator function for A, equals 1 if A is true and 0 otherwise
n	Number of participants
d	$\lvert G\rvert = d$, number of groups
G	Set of groups
ε	Privacy parameter
(g_i, v_i)	$g_i \in G, v_i \in [-1, 1]$, group and value of participant i
(g'_i, v'_i)	Perturbed group and value response of participant i, $(g'_i, v'_i) = M(g_i, v_i)$
k	Discretization parameter for the Group NPRR method
n_g	$n_g = \sum_i^n \mathbf{1}_{g_i=g}$, number of participants in group g
s_g	$s_g = \sum_i^n \mathbf{1}_{g_i=g} \cdot v_i$, sum of values of group g
m_g	$m_g = \frac{1}{n_g} \sum_i^n \mathbf{1}_{g_i=g} \cdot v_i$, mean of group g
$\hat{n}_g$	Estimated number of participants in group g
$\hat{m}_g$	Estimated mean of group g
$\hat{s}_g$	Estimated sum of group g
$\mathrm{Lap}(\lambda)$	Sample from a Laplace distribution with scale λ and probability density function $f(x) = \frac{1}{2\lambda} e^{-\frac{\lvert x\rvert}{\lambda}}$
$\mathrm{Bernoulli}(p)$	Sample from a Bernoulli distribution with parameter p.

Table 7: Categories from MIMIC-IV groups and their corresponding indices used in the figures in this paper.

Group Name	Categories
Admission Type	AMBULATORY OBSERVATION (0), DIRECT EMER. (1), DIRECT OBSERVATION (2), ELECTIVE (3), EU OBSERVATION (4), EW EMER. (5), OBSERVATION ADMIT (6), SURGICAL SAME DAY ADMISSION (7), URGENT (8)
Admission Location	AMBULATORY SURGERY TRANSFER (0), CLINIC REFERRAL (1), EMERGENCY ROOM (2), INFORMATION NOT AVAILABLE (3), INTERNAL TRANSFER TO OR FROM PSYCH (4), PACU (5), PHYSICIAN REFERRAL (6), PROCEDURE SITE (7), TRANSFER FROM HOSPITAL (8), TRANSFER FROM SKILLED NURSING FACILITY (9), WALK-IN/SELF REFERRAL (10)
Discharge Location	ACUTE HOSPITAL (0), AGAINST ADVICE (1), ASSISTED LIVING (2), CHRONIC/LONG TERM ACUTE CARE (3), DIED (4), HEALTHCARE FACILITY (5), HOME (6), HOME HEALTH CARE (7), HOSPICE (8), OTHER FACILITY (9), PSYCH FACILITY (10), REHAB (11), SKILLED NURSING FACILITY (12)

Table 8: Mean scaled absolute error (absolute error divided by the data range of the dataset) of the Piecewise mechanism for different settings of ε_1, ε and $\varepsilon_2 = \varepsilon - \varepsilon_1$ for 2 groups averaged over the four synthetic datasets. Each group consists of 10000 participants. The lowest error for each ε is shown in bold. The standard deviation is shown in parentheses below the mean error.

$\varepsilon_1/\varepsilon$ / ε	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	7.51×10^{1}	1.55×10^{1}	1.71	$\mathbf{3.91\times10^{-1}}$	4.06×10^{-1}	4.67×10^{-1}	6.13×10^{-1}	9.29×10^{-1}	1.80
	(1.92×10^{3})	(3.84×10^{2})	(6.67×10^{1})	(1.21)	(5.47×10^{-1})	(4.03×10^{-1})	(5.05×10^{-1})	(7.36×10^{-1})	(1.38)
0.5	5.71×10^{-2}	$\mathbf{4.56\times10^{-2}}$	4.90×10^{-2}	5.38×10^{-2}	6.39×10^{-2}	7.80×10^{-2}	1.04×10^{-1}	1.57×10^{-1}	3.01×10^{-1}
	(1.06×10^{-1})	(3.64×10^{-2})	(3.72×10^{-2})	(4.11×10^{-2})	(4.84×10^{-2})	(5.95×10^{-2})	(7.93×10^{-2})	(1.15×10^{-1})	(2.28×10^{-1})
1.0	2.38×10^{-2}	$\mathbf{2.04\times10^{-2}}$	2.18×10^{-2}	2.39×10^{-2}	2.74×10^{-2}	3.33×10^{-2}	4.34×10^{-2}	6.41×10^{-2}	1.28×10^{-1}
	(2.00×10^{-2})	(1.55×10^{-2})	(1.64×10^{-2})	(1.82×10^{-2})	(2.09×10^{-2})	(2.54×10^{-2})	(3.35×10^{-2})	(4.91×10^{-2})	(9.62×10^{-2})
2.0	1.04×10^{-2}	$\mathbf{8.68\times10^{-3}}$	8.79×10^{-3}	9.53×10^{-3}	1.13×10^{-2}	1.37×10^{-2}	1.79×10^{-2}	2.58×10^{-2}	5.20×10^{-2}
	(8.49×10^{-3})	(6.55×10^{-3})	(6.57×10^{-3})	(7.27×10^{-3})	(8.35×10^{-3})	(1.02×10^{-2})	(1.36×10^{-2})	(1.96×10^{-2})	(3.94×10^{-2})
4.0	4.68×10^{-3}	3.43×10^{-3}	$\mathbf{3.37\times10^{-3}}$	3.58×10^{-3}	4.28×10^{-3}	5.40×10^{-3}	7.04×10^{-3}	1.09×10^{-2}	2.28×10^{-2}
	(3.76×10^{-3})	(2.79×10^{-3})	(2.67×10^{-3})	(2.77×10^{-3})	(3.29×10^{-3})	(4.07×10^{-3})	(5.37×10^{-3})	(8.03×10^{-3})	(1.71×10^{-2})
6.0	2.97×10^{-3}	2.03×10^{-3}	$\mathbf{1.92\times10^{-3}}$	2.03×10^{-3}	2.31×10^{-3}	3.06×10^{-3}	4.30×10^{-3}	6.72×10^{-3}	1.45×10^{-2}
	(2.42×10^{-3})	(1.71×10^{-3})	(1.57×10^{-3})	(1.62×10^{-3})	(1.81×10^{-3})	(2.37×10^{-3})	(3.28×10^{-3})	(5.10×10^{-3})	(1.10×10^{-2})
8.0	2.22×10^{-3}	1.40×10^{-3}	$\mathbf{1.21\times10^{-3}}$	1.28×10^{-3}	1.58×10^{-3}	2.07×10^{-3}	2.98×10^{-3}	4.72×10^{-3}	1.05×10^{-2}
	(1.81×10^{-3})	(1.19×10^{-3})	(9.72×10^{-4})	(1.04×10^{-3})	(1.27×10^{-3})	(1.63×10^{-3})	(2.32×10^{-3})	(3.60×10^{-3})	(8.07×10^{-3})
10.0	1.79×10^{-3}	1.03×10^{-3}	$\mathbf{8.21\times10^{-4}}$	8.72×10^{-4}	1.11×10^{-3}	1.55×10^{-3}	2.21×10^{-3}	3.69×10^{-3}	8.15×10^{-3}
	(1.50×10^{-3})	(9.25×10^{-4})	(7.02×10^{-4})	(7.12×10^{-4})	(8.73×10^{-4})	(1.22×10^{-3})	(1.72×10^{-3})	(2.81×10^{-3})	(6.22×10^{-3})

Table 9: Mean scaled absolute error (absolute error divided by the data range of the dataset) of the Piecewise mechanism for different settings of ε_1, ε and $\varepsilon_2 = \varepsilon - \varepsilon_1$ for 8 groups averaged over the four synthetic datasets. Each group consists of 10000 participants. The lowest error for each ε is shown in bold. The standard deviation is shown in parentheses below the mean error.

$\varepsilon_1/\varepsilon$ / ε	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	$\mathbf{6.66\times10^{-1}}$	1.42	4.13	3.43	4.48	5.86	8.49	3.48×10^{1}	2.33×10^{1}
	(2.78)	(8.82)	(4.44×10^{1})	(2.16×10^{1})	(2.73×10^{1})	(3.40×10^{1})	(6.48×10^{1})	(6.28×10^{2})	(1.56×10^{2})
0.5	7.01×10^{-1}	8.89×10^{-1}	5.11×10^{-1}	3.17×10^{-1}	$\mathbf{2.61\times10^{-1}}$	2.91×10^{-1}	3.70×10^{-1}	5.33×10^{-1}	1.01
	(3.51)	(8.59)	(5.33)	(3.98)	(2.65×10^{-1})	(2.43×10^{-1})	(2.98×10^{-1})	(4.19×10^{-1})	(7.95×10^{-1})
1.0	5.54×10^{-1}	1.30×10^{-1}	8.72×10^{-2}	$\mathbf{8.56\times10^{-2}}$	9.28×10^{-2}	1.07×10^{-1}	1.33×10^{-1}	1.85×10^{-1}	3.55×10^{-1}
	(4.48)	(1.06)	(8.06×10^{-2})	(6.85×10^{-2})	(7.19×10^{-2})	(8.19×10^{-2})	(1.01×10^{-1})	(1.42×10^{-1})	(2.68×10^{-1})
2.0	1.14×10^{-1}	3.72×10^{-2}	3.02×10^{-2}	$\mathbf{2.88\times10^{-2}}$	2.97×10^{-2}	3.22×10^{-2}	3.87×10^{-2}	5.24×10^{-2}	9.67×10^{-2}
	(2.18)	(3.53×10^{-2})	(2.45×10^{-2})	(2.21×10^{-2})	(2.27×10^{-2})	(2.45×10^{-2})	(2.90×10^{-2})	(3.94×10^{-2})	(7.28×10^{-2})
4.0	2.57×10^{-2}	1.30×10^{-2}	9.49×10^{-3}	8.03×10^{-3}	$\mathbf{7.55\times10^{-3}}$	8.07×10^{-3}	9.72×10^{-3}	1.36×10^{-2}	2.62×10^{-2}
	(3.04×10^{-2})	(1.21×10^{-2})	(7.75×10^{-3})	(6.28×10^{-3})	(5.91×10^{-3})	(6.16×10^{-3})	(7.38×10^{-3})	(1.03×10^{-2})	(2.00×10^{-2})
6.0	1.51×10^{-2}	7.04×10^{-3}	4.54×10^{-3}	3.63×10^{-3}	$\mathbf{3.38\times10^{-3}}$	3.73×10^{-3}	4.77×10^{-3}	7.23×10^{-3}	1.49×10^{-2}
	(1.69×10^{-2})	(6.42×10^{-3})	(3.89×10^{-3})	(2.98×10^{-3})	(2.65×10^{-3})	(2.94×10^{-3})	(3.69×10^{-3})	(5.48×10^{-3})	(1.14×10^{-2})
8.0	1.03×10^{-2}	4.44×10^{-3}	2.72×10^{-3}	2.05×10^{-3}	$\mathbf{1.94\times10^{-3}}$	2.29×10^{-3}	3.17×10^{-3}	5.02×10^{-3}	1.06×10^{-2}
	(1.10×10^{-2})	(4.07×10^{-3})	(2.40×10^{-3})	(1.72×10^{-3})	(1.56×10^{-3})	(1.80×10^{-3})	(2.47×10^{-3})	(3.82×10^{-3})	(8.13×10^{-3})
10.0	7.80×10^{-3}	3.17×10^{-3}	1.80×10^{-3}	1.29×10^{-3}	$\mathbf{1.28\times10^{-3}}$	1.63×10^{-3}	2.33×10^{-3}	3.84×10^{-3}	8.39×10^{-3}
	(8.14×10^{-3})	(2.92×10^{-3})	(1.63×10^{-3})	(1.10×10^{-3})	(1.04×10^{-3})	(1.28×10^{-3})	(1.81×10^{-3})	(2.94×10^{-3})	(6.34×10^{-3})

Table 10: Mean scaled absolute error (absolute error divided by the data range of the dataset) of the Piecewise mechanism for different settings of ε_1, ε and $\varepsilon_2 = \varepsilon - \varepsilon_1$ for 64 groups averaged over the four synthetic datasets. Each group consists of 10000 participants. The lowest error for each ε is shown in bold. The standard deviation is shown in parentheses below the mean error.

$\varepsilon_1/\varepsilon$ / ε	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	**6.21×10^{-1}**	1.59	2.37	3.20	7.60×10^{1}	7.37	1.15×10^{1}	4.28×10^{2}	4.35×10^{1}
	(2.48)	(1.17×10^{1})	(1.42×10^{1})	(1.55×10^{1})	(1.44×10^{3})	(3.65×10^{1})	(6.14×10^{1})	(8.28×10^{3})	(2.32×10^{2})
0.5	8.62	**1.18**	2.08	2.64	4.54	5.50	1.70×10^{1}	1.41×10^{1}	1.17×10^{2}
	(1.59×10^{2})	(5.40)	(1.29×10^{1})	(1.20×10^{1})	(3.21×10^{1})	(2.82×10^{1})	(2.33×10^{2})	(9.37×10^{1})	(2.39×10^{3})
1.0	**5.64×10^{-1}**	1.04	1.54	2.17	2.40	3.25	4.42	4.08	1.53×10^{1}
	(2.29)	(4.65)	(7.77)	(1.40×10^{1})	(1.41×10^{1})	(3.16×10^{1})	(7.99×10^{1})	(6.63×10^{1})	(6.48×10^{2})
2.0	5.48×10^{-1}	9.50×10^{-1}	1.07	8.49×10^{-1}	3.31×10^{-1}	**2.37×10^{-1}**	2.44×10^{-1}	3.05×10^{-1}	5.17×10^{-1}
	(2.22)	(6.04)	(1.09×10^{1})	(1.57×10^{1})	(2.71)	(6.10×10^{-1})	(2.04×10^{-1})	(2.42×10^{-1})	(4.00×10^{-1})
4.0	6.98×10^{-1}	5.49×10^{-1}	8.13×10^{-2}	4.86×10^{-2}	3.78×10^{-2}	3.27×10^{-2}	**3.20×10^{-2}**	3.69×10^{-2}	5.94×10^{-2}
	(4.66)	(1.12×10^{1})	(7.21×10^{-1})	(4.49×10^{-2})	(3.07×10^{-2})	(2.53×10^{-2})	(2.44×10^{-2})	(2.79×10^{-2})	(4.49×10^{-2})
6.0	7.50×10^{-1}	6.54×10^{-2}	2.72×10^{-2}	1.64×10^{-2}	1.12×10^{-2}	8.89×10^{-3}	**8.55×10^{-3}**	1.03×10^{-2}	1.86×10^{-2}
	(8.55)	(5.66×10^{-1})	(2.81×10^{-2})	(1.47×10^{-2})	(9.19×10^{-3})	(6.96×10^{-3})	(6.56×10^{-3})	(7.85×10^{-3})	(1.41×10^{-2})
8.0	4.76×10^{-1}	3.05×10^{-2}	1.35×10^{-2}	7.26×10^{-3}	4.65×10^{-3}	**3.71×10^{-3}**	3.93×10^{-3}	5.50×10^{-3}	1.12×10^{-2}
	(1.03×10^{1})	(3.81×10^{-2})	(1.35×10^{-2})	(6.53×10^{-3})	(3.92×10^{-3})	(2.97×10^{-3})	(3.05×10^{-3})	(4.22×10^{-3})	(8.48×10^{-3})
10.0	1.52×10^{-1}	1.87×10^{-2}	7.59×10^{-3}	3.83×10^{-3}	2.42×10^{-3}	**2.08×10^{-3}**	2.53×10^{-3}	3.92×10^{-3}	8.46×10^{-3}
	(2.27)	(2.12×10^{-2})	(7.44×10^{-3})	(3.47×10^{-3})	(2.06×10^{-3})	(1.68×10^{-3})	(1.97×10^{-3})	(3.01×10^{-3})	(6.42×10^{-3})

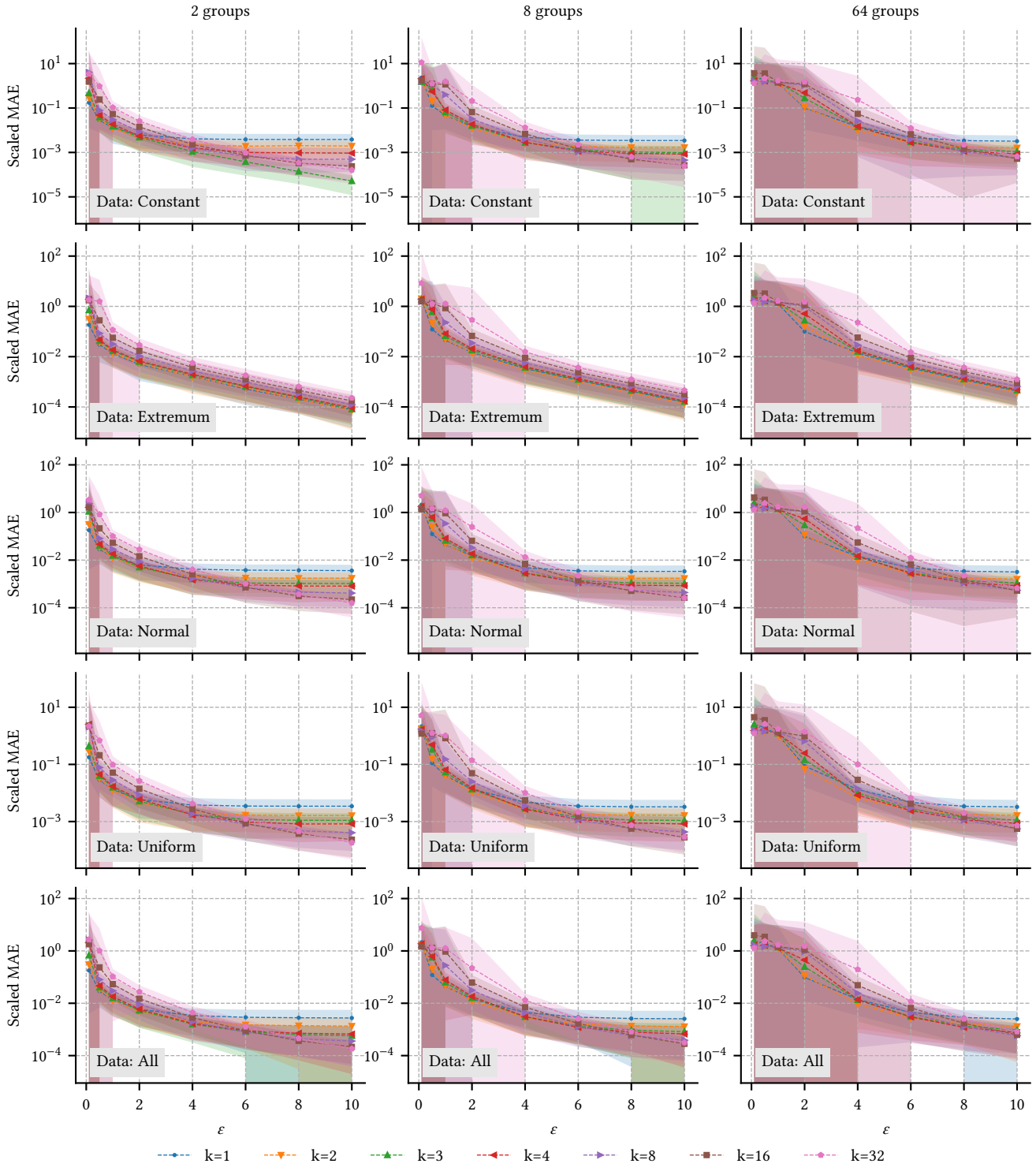


Figure 7: Scaled absolute error (absolute error divided by the data range of the dataset) for the Group NPRR method with different settings of k. The group size in all three subfigures is equal to 10 000. Note that this also means, that all three have a different number of total participants: 20 000 for 2 groups, 80 000 for 8 groups, and 64 000 for 64 groups. Shaded regions indicate the standard deviation over the randomness of the mechanism.

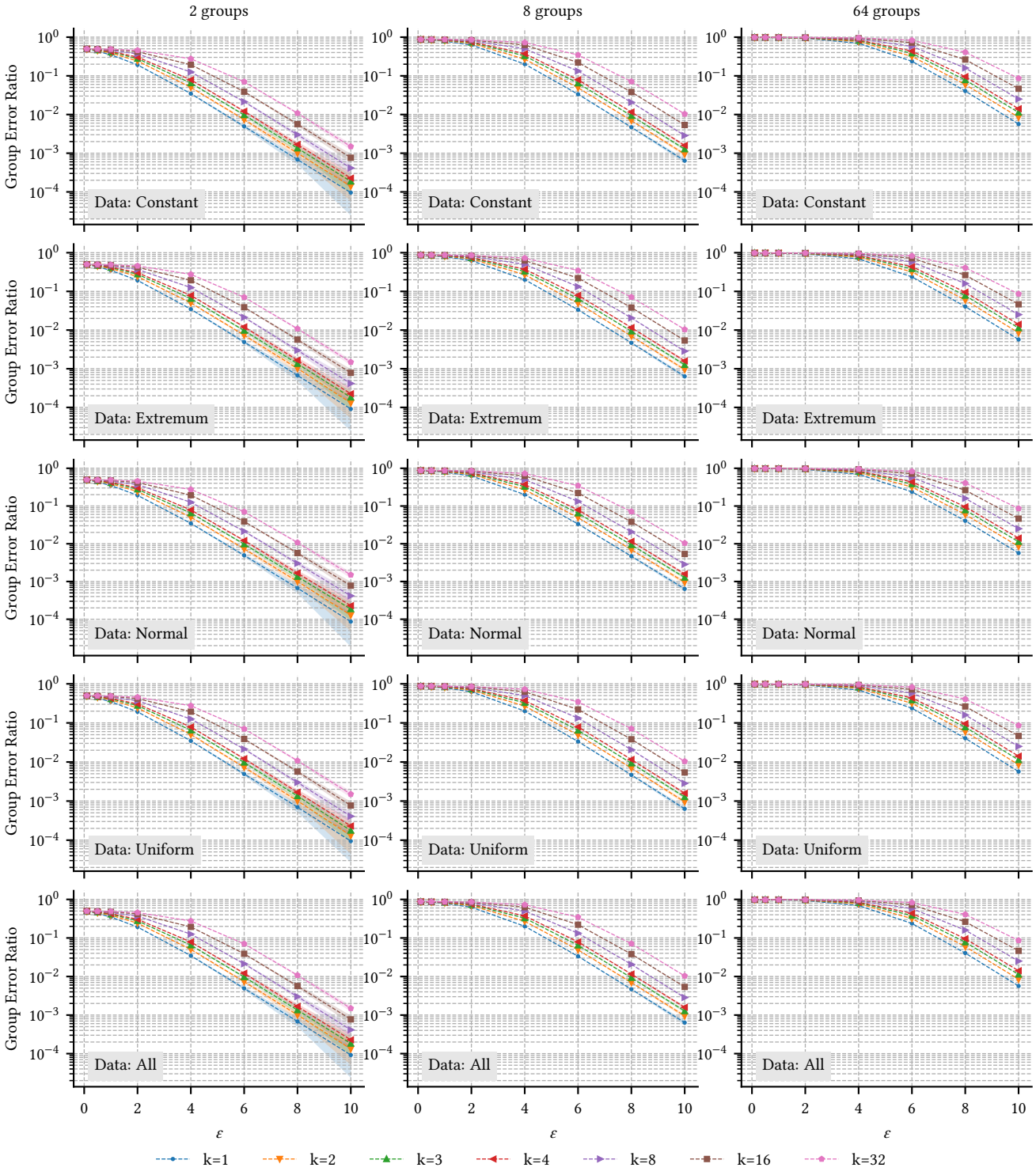


Figure 8: Group error ratio for the Group NPRR method with different settings of k. The group error ratio gives the number of incorrect group responses divided by the number of participants. This figure uses the same settings as Figure 7.

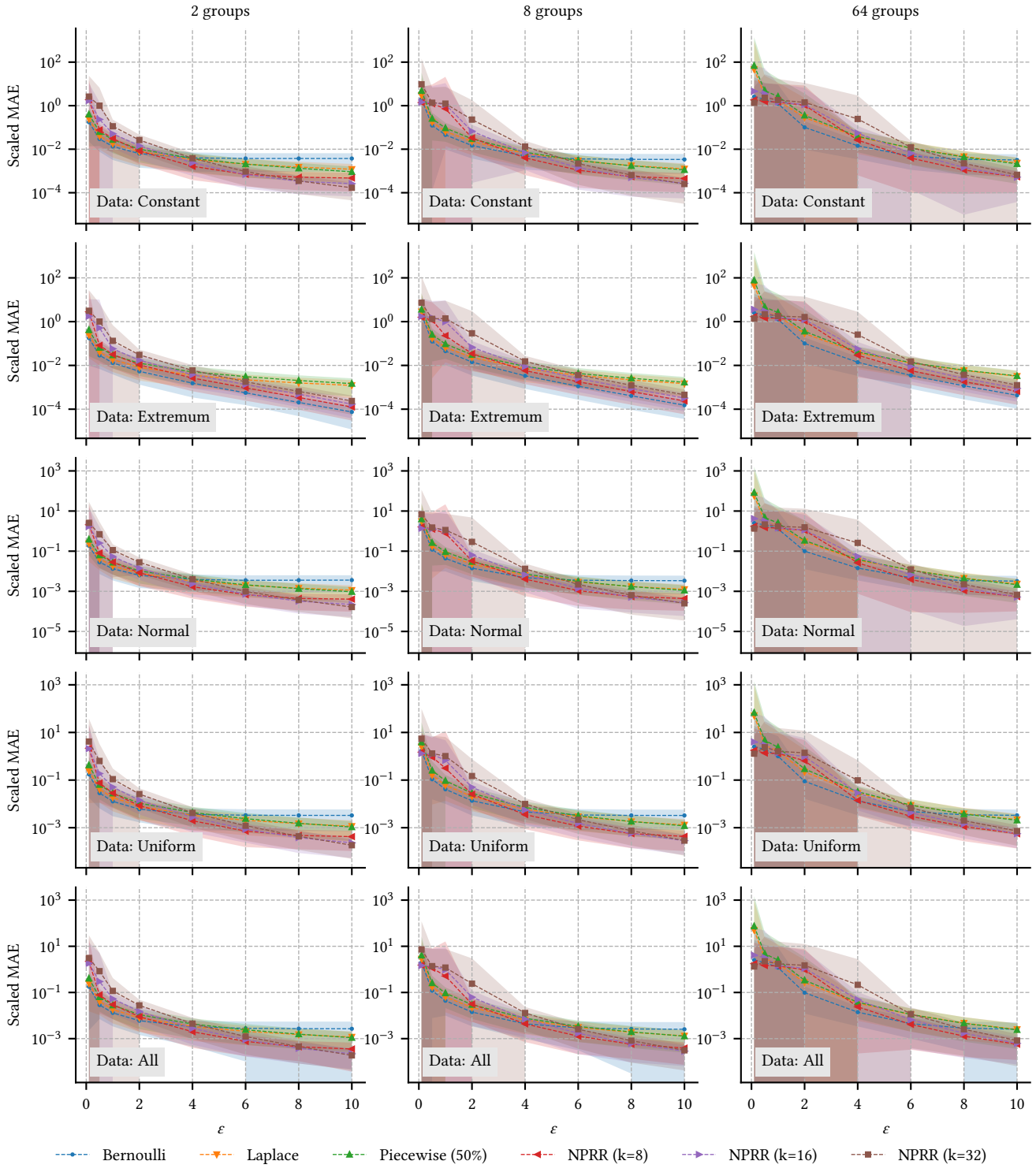


Figure 9: Scaled absolute error (absolute error divided by the data range of the dataset) for the proposed methods. The group size in all three subfigures is equal to 10 000. Note that this also means, that all three have a different number of total participants: 20 000 for 2 groups, 80 000 for 8 groups, and 64 000 for 64 groups. Shaded regions indicate the standard deviation over the randomness of the mechanism.

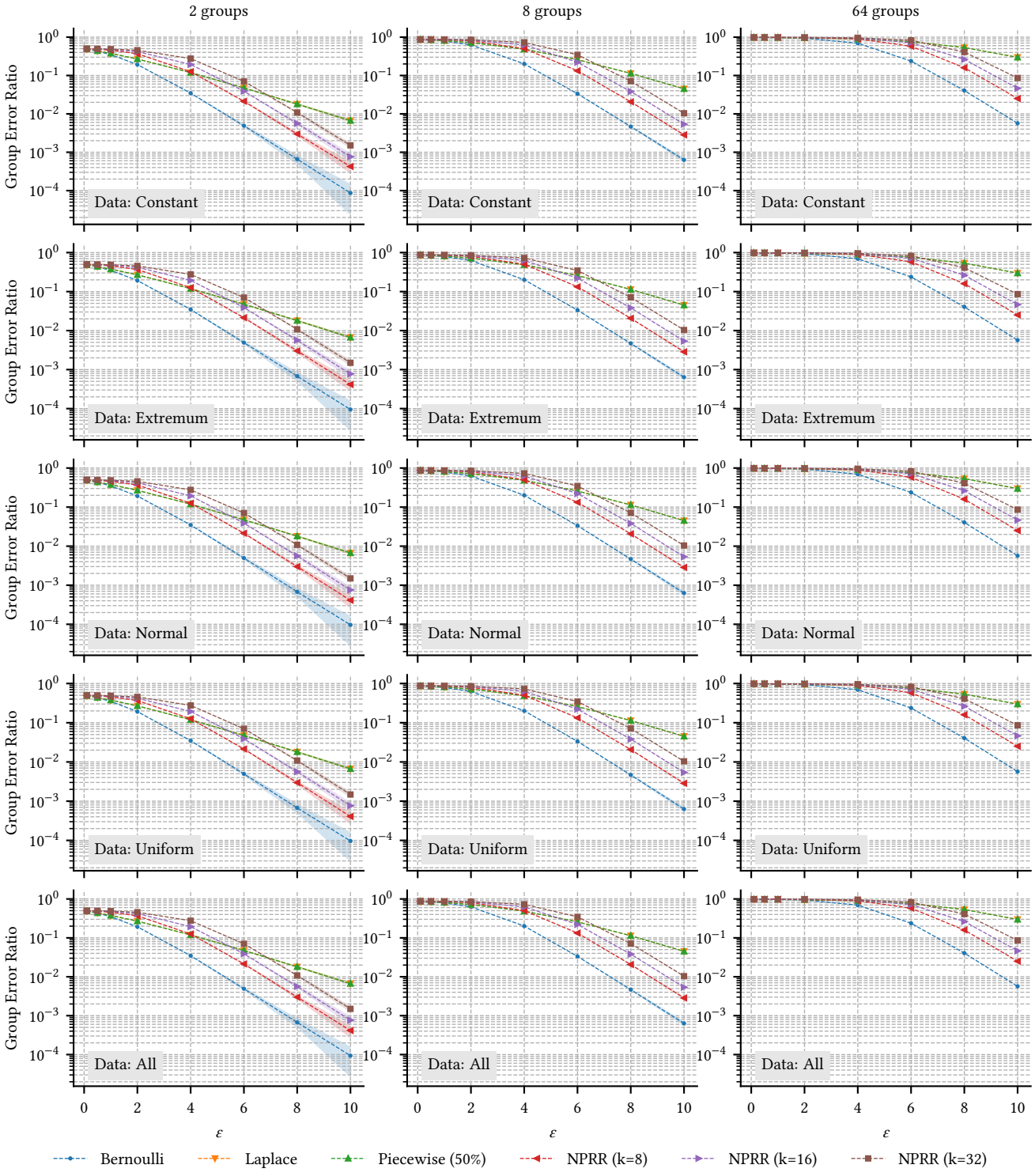


Figure 10: Group error ratio for the proposed methods. The group error ratio gives the number of incorrect group responses divided by the number of participants. This figure uses the same settings as Figure 9.

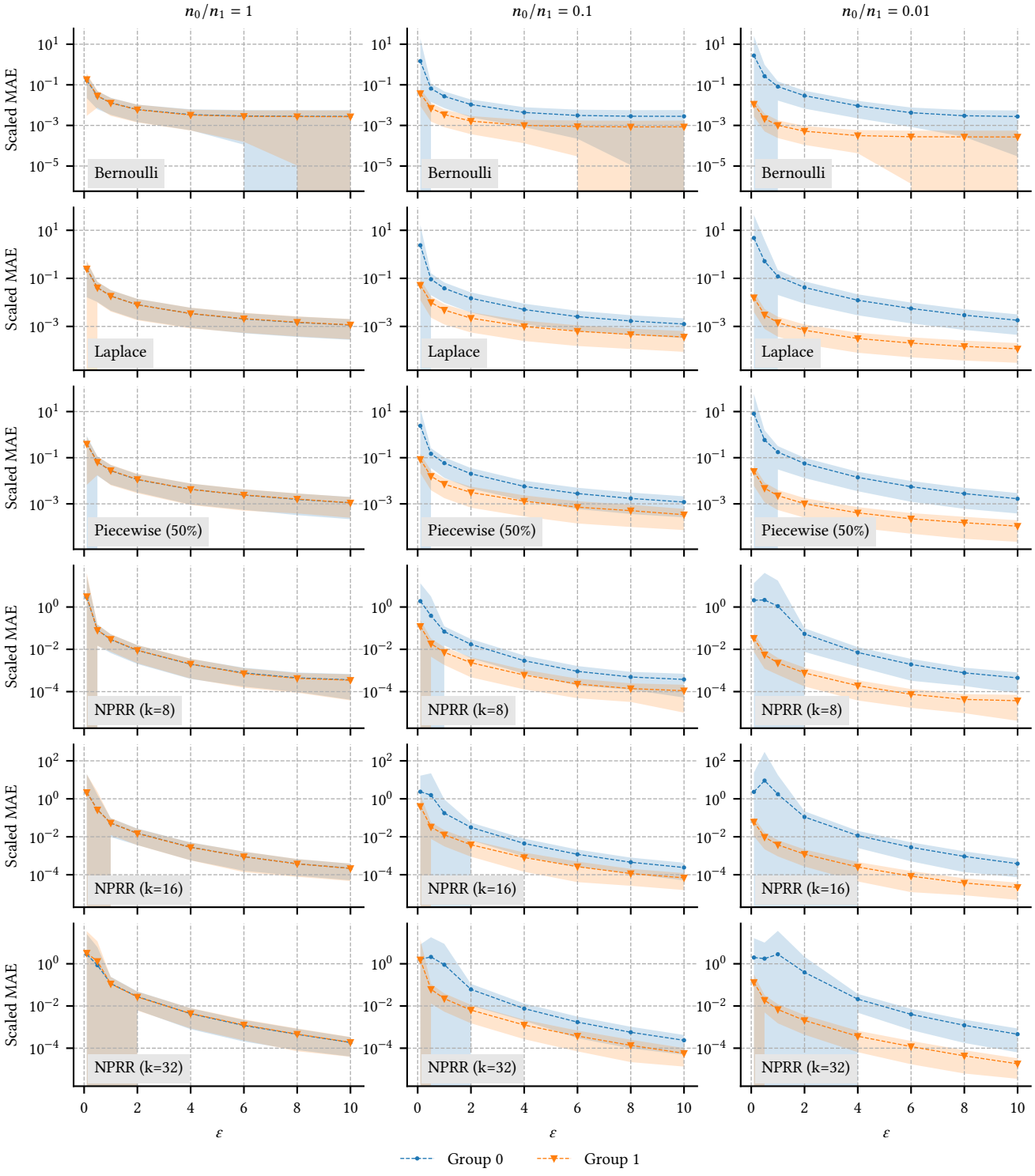
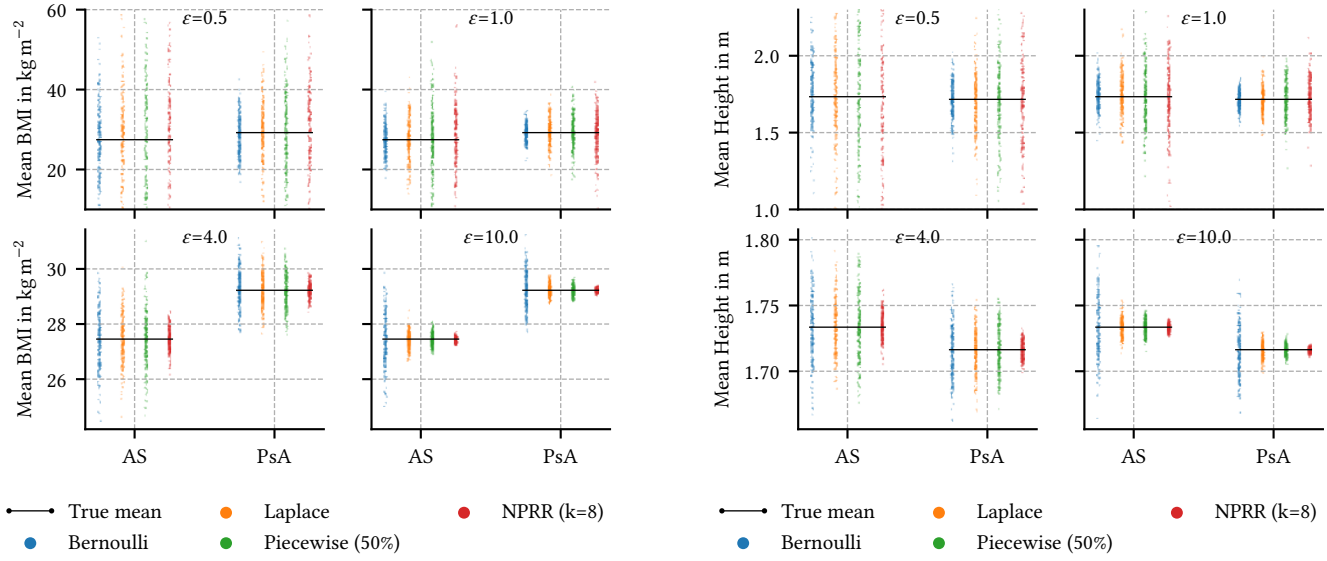
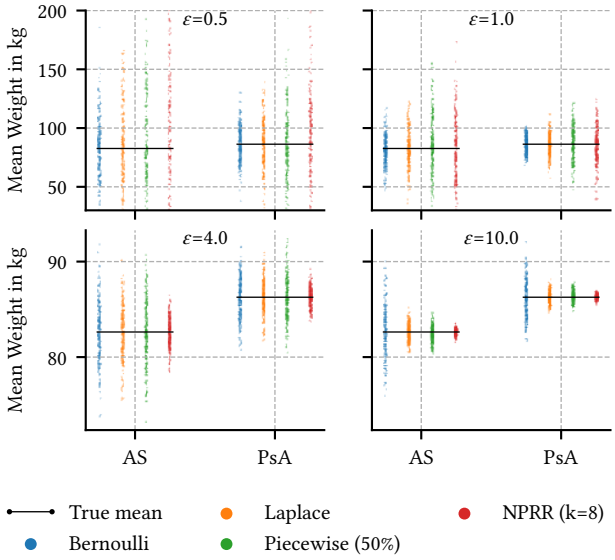


Figure 11: Scaled absolute error (absolute error divided by the data range of the dataset) for the different proposed methods with different group sizes and imbalances. The error is shown per method and group. Shaded regions indicate the standard deviation over the randomness of the mechanism.

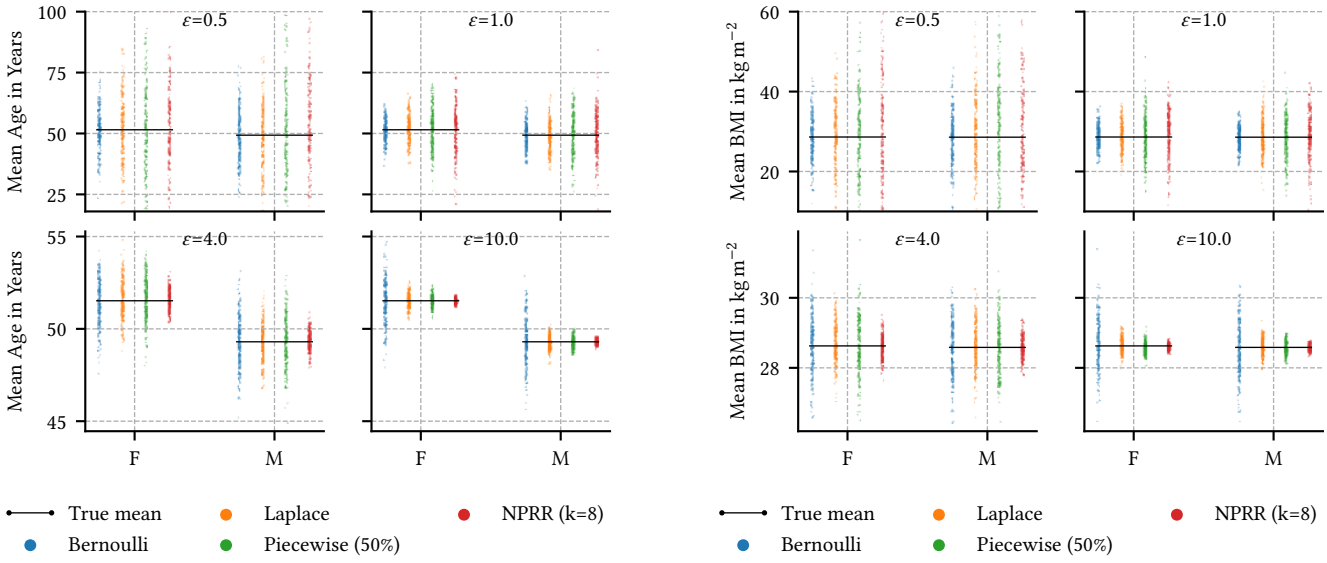


(a) Value: BMI **(b) Value: Height**



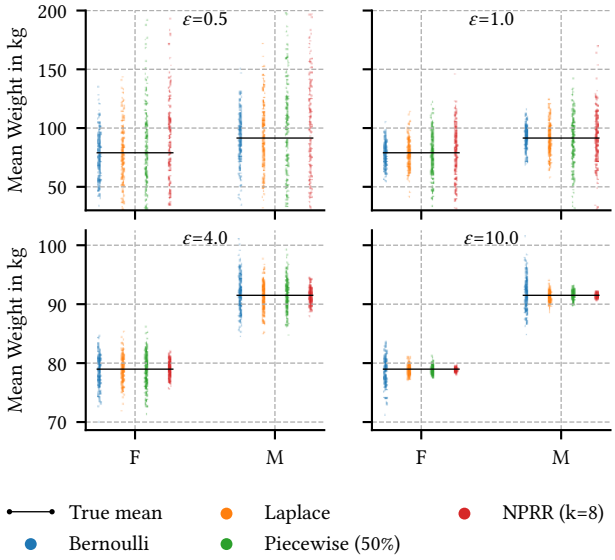
(c) Value: Weight

Figure 12: Individual results for the mean estimation for different groups and values using data from the AQUILA dataset with disease as the group. Participants per group: AS: 683, **PsA:** 1278. **Each method was executed** 200 **times and each resulting mean estimate is shown as a single point to visualize the randomness of the mechanism. The true mean is shown as a solid line.**



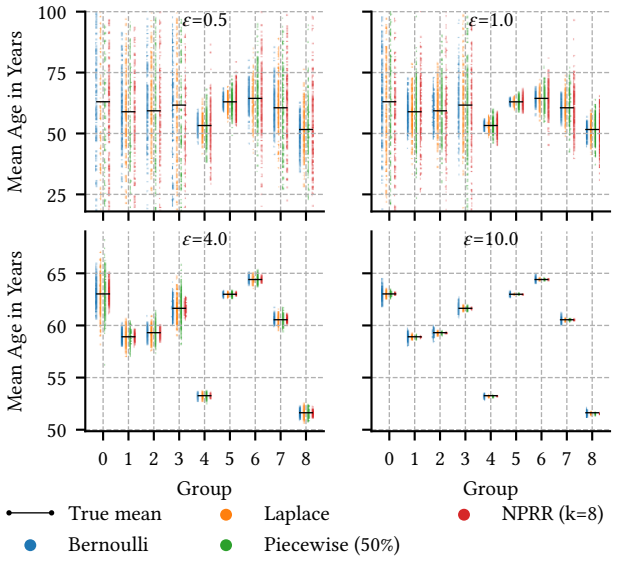
(a) Value: Age

(b) Value: BMI

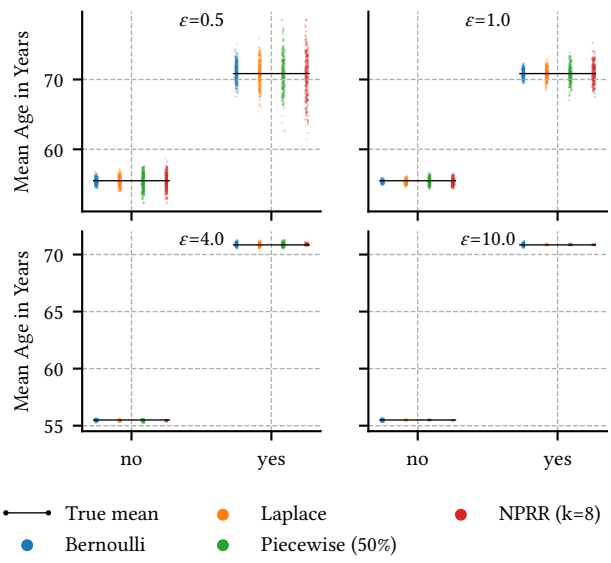


(c) Value: Weight

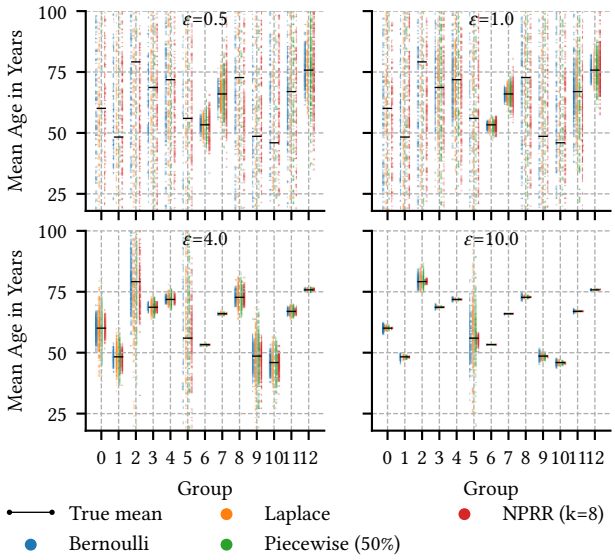
Figure 13: Individual results for the mean estimation for different groups and values using data from the AQUILA dataset with gender as the group. Participants per group: F: 993, **M:** 921**. Each method was executed** 200 **times and each resulting mean estimate is shown as a single point to visualize the randomness of the mechanism. The true mean is shown as a solid line.**



(a) Group: Admission Type. Group Sizes: 0: 6620, **1:** 19 519, **2:** 18 682, **3:** 10 557, **4:** 94 739, **5:** 149 278, **6:** 52 641, **7:** 34 184, **8:** 44 642

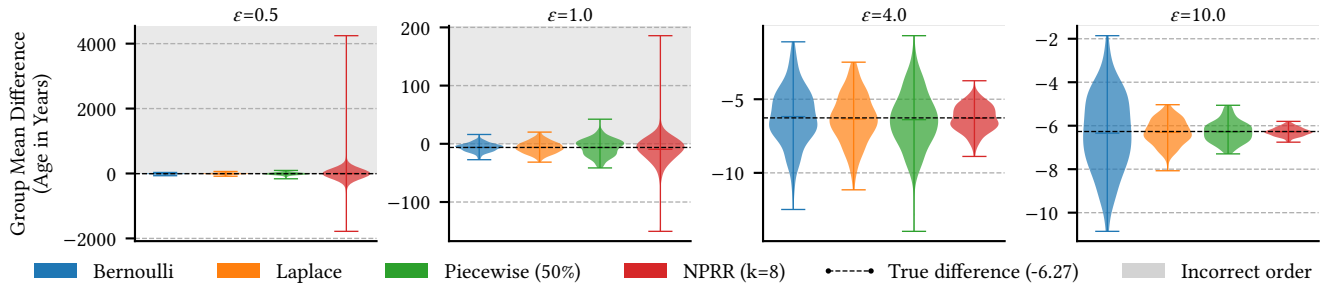


(b) Group: Deceased. Group sizes: no: 324 785, **yes:** 106 077

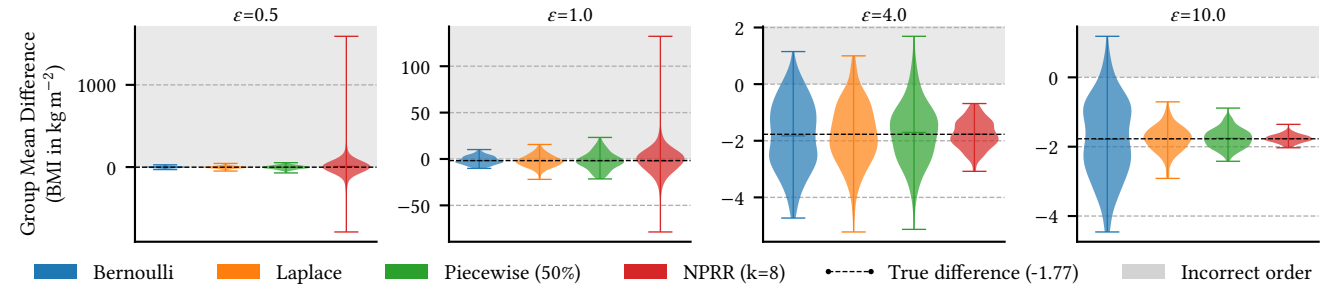


(c) Group: Discharge Location. Group Sizes: 0: 1610, **1:** 2587, **2:** 550, **3:** 7144, **4:** 8506, **5:** 42, **6:** 155 321, **7:** 75 545, **8:** 3469, **9:** 1354, **10:** 2262, **11:** 10 521, **12:** 43 013

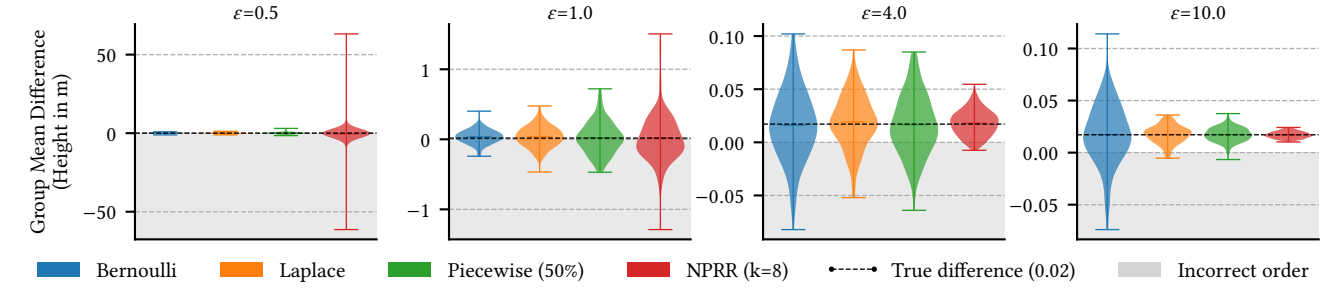
Figure 14: Individual results for the mean estimation for different groups and values using data from the MIMIC-IV dataset with age as the value. Each method was executed 200 **times and each resulting mean estimate is shown as a single point to visualize the randomness of the mechanism. The true mean is shown as a solid line.**



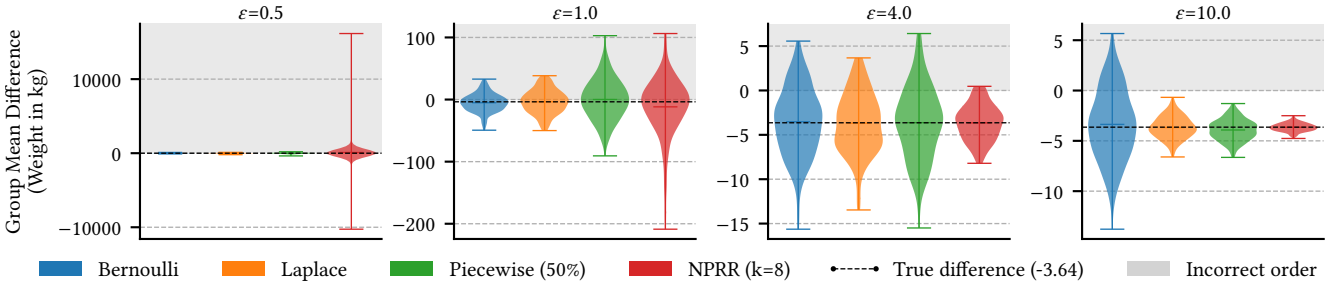
(a) Value: Age



(b) Value: BMI

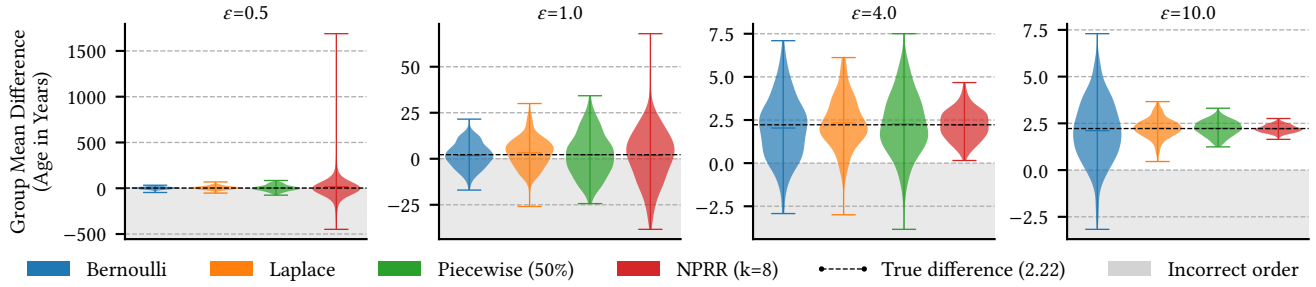


(c) Value: Height

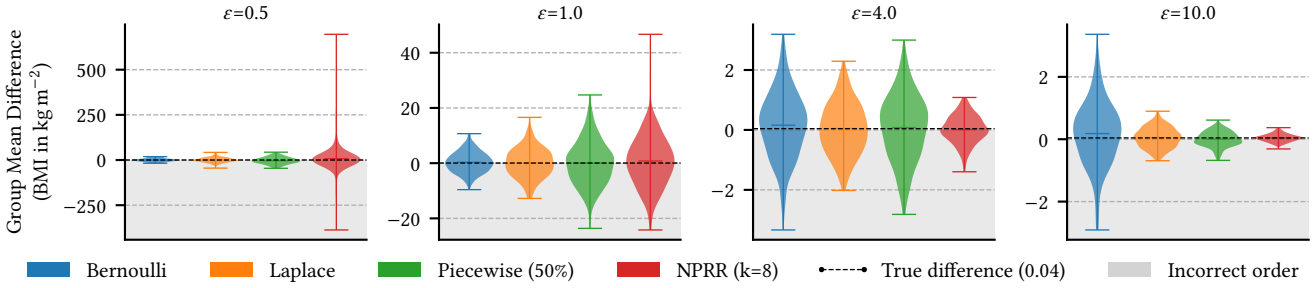


(d) Value: Weight

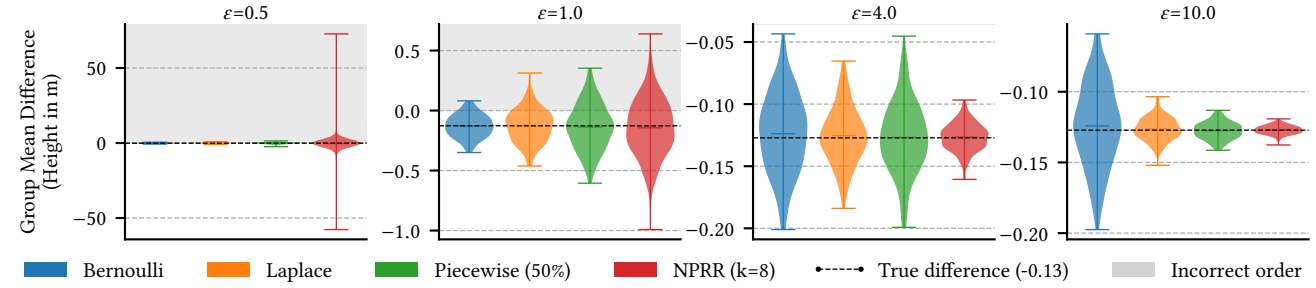
Figure 15: Individual results for the mean estimation for different groups and values using data from the AQUILA dataset with disease as the group. Participants per group: AS: 683, PsA: 1278. Each method was executed 200 times. Each data point shows the percentage of runs, where the order of the mean estimate for the two groups was correct (i.e., the same as for the true group means). The difference between the true means is displayed in the figure.



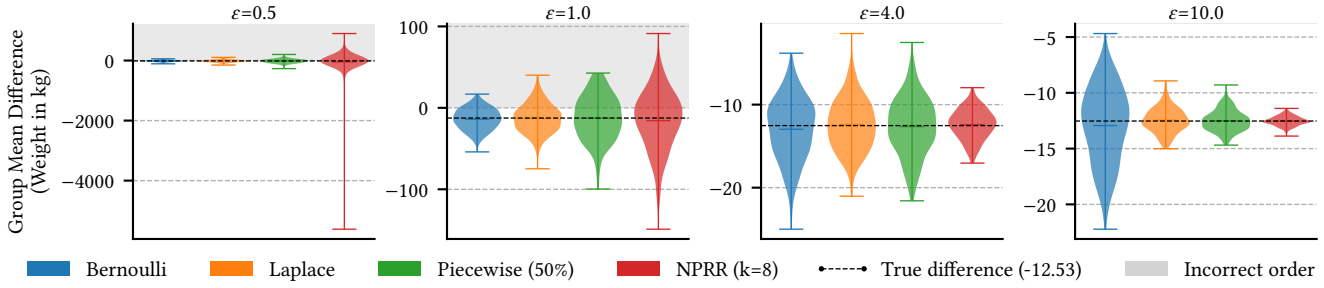
(a) Value: Age



(b) Value: BMI

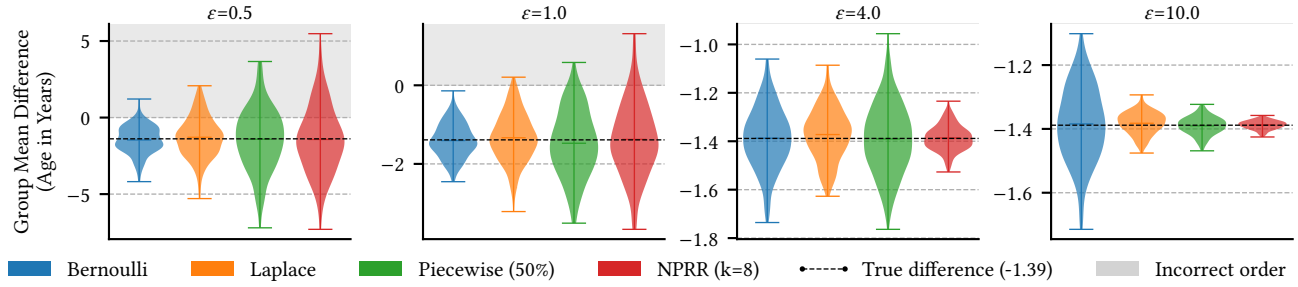


(c) Value: Height

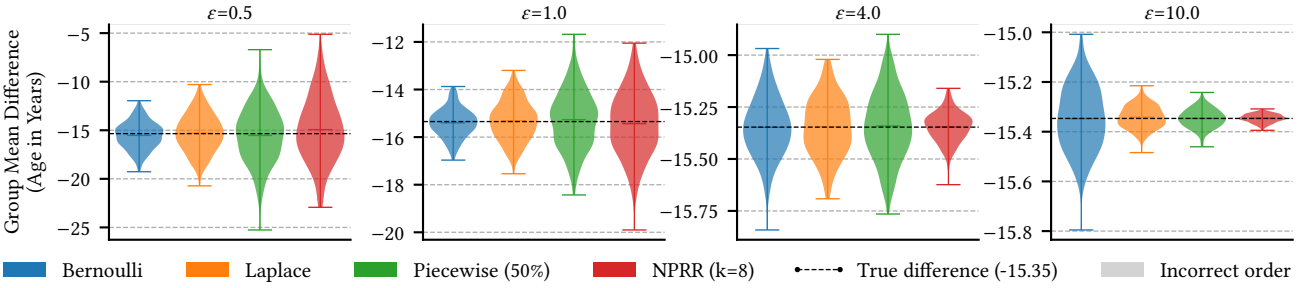


(d) Value: Weight

Figure 16: Individual results for the mean estimation for different groups and values using data from the AQUILA dataset with gender as the group. Participants per group: F: 993, M: 921. Each method was executed 200 times. Each data point shows the percentage of runs, where the order of the mean estimate for the two groups was correct (i.e., the same as for the true group means). The difference between the true means is displayed in the figure.

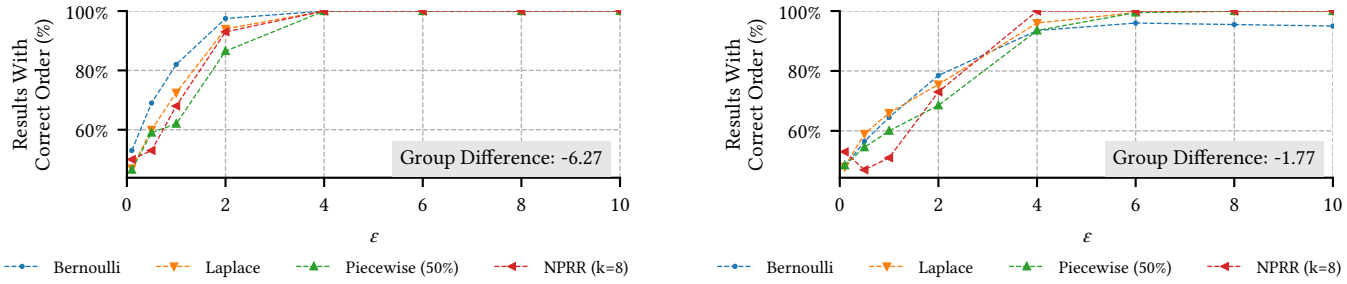


(a) Group: Gender. Group sizes: F: 224 796, **M:** 206 066

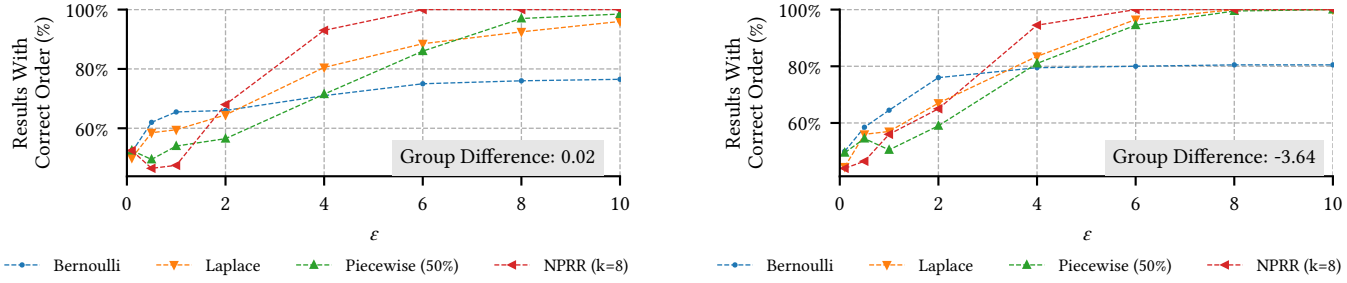


(b) Group: Deceased. Group sizes: no: 324 785, **yes:** 106 077

Figure 17: Individual results for the mean estimation for different groups and values using data from the MIMIC-IV dataset with age as the value. Each method was executed 200 **times.**

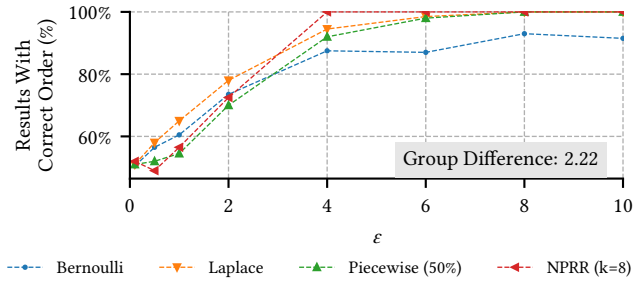


(a) Value: Age **(b) Value: BMI**

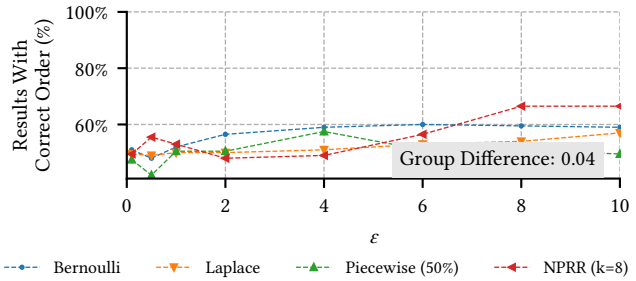


(c) Value: Height **(d) Value: Weight**

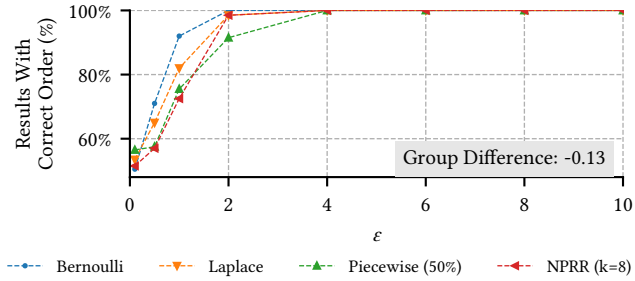
Figure 18: Individual results for the mean estimation for different groups and values using data from the AQUILA dataset with disease as the group. Participants per group: AS: 683, **PsA:** 1278. **Each method was executed** 200 **times. Each data point shows the percentage of runs, where the order of the mean estimate for the two groups was correct (i.e., the same as for the true group means). The difference between the true means is displayed in the figure.**



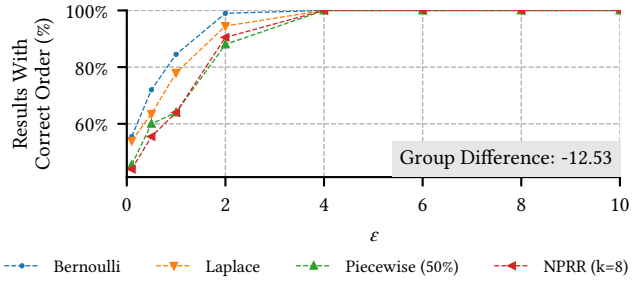
(a) Value: Age



(b) Value: BMI

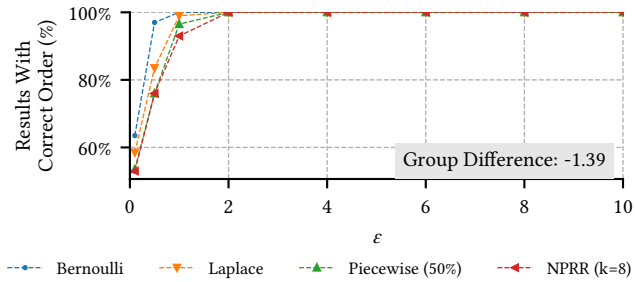


(c) Value: Height

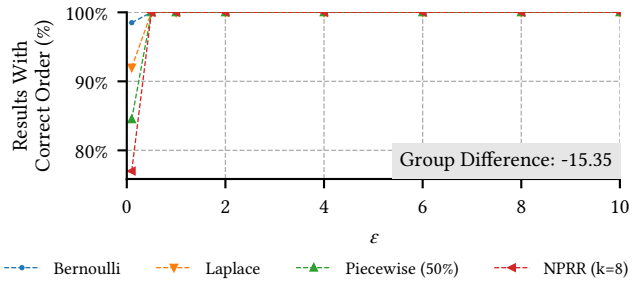


(d) Value: Weight

Figure 19: Individual results for the mean estimation for different groups and values using data from the AQUILA dataset with gender as the group. Participants per group: F: 993, **M:** 921**. Each method was executed** 200 **times. Each data point shows the percentage of runs, where the order of the mean estimate for the two groups was correct (i.e., the same as for the true group means). The difference between the true means is displayed in the figure.**



(a) Group: Gender. Group sizes: F: 224 796, **M:** 206 066



(b) Group: Deceased. Group sizes: no: 324 785, **yes:** 106 077

Figure 20: Individual results for the mean estimation for different groups and values using data from the MIMIC-IV dataset with age as the value. Each method was executed 200 **times. Each data point shows the percentage of runs, where the order of the mean estimate for the two groups was correct (i.e., the same as for the true group means). The difference between the true means is displayed in the figure.**

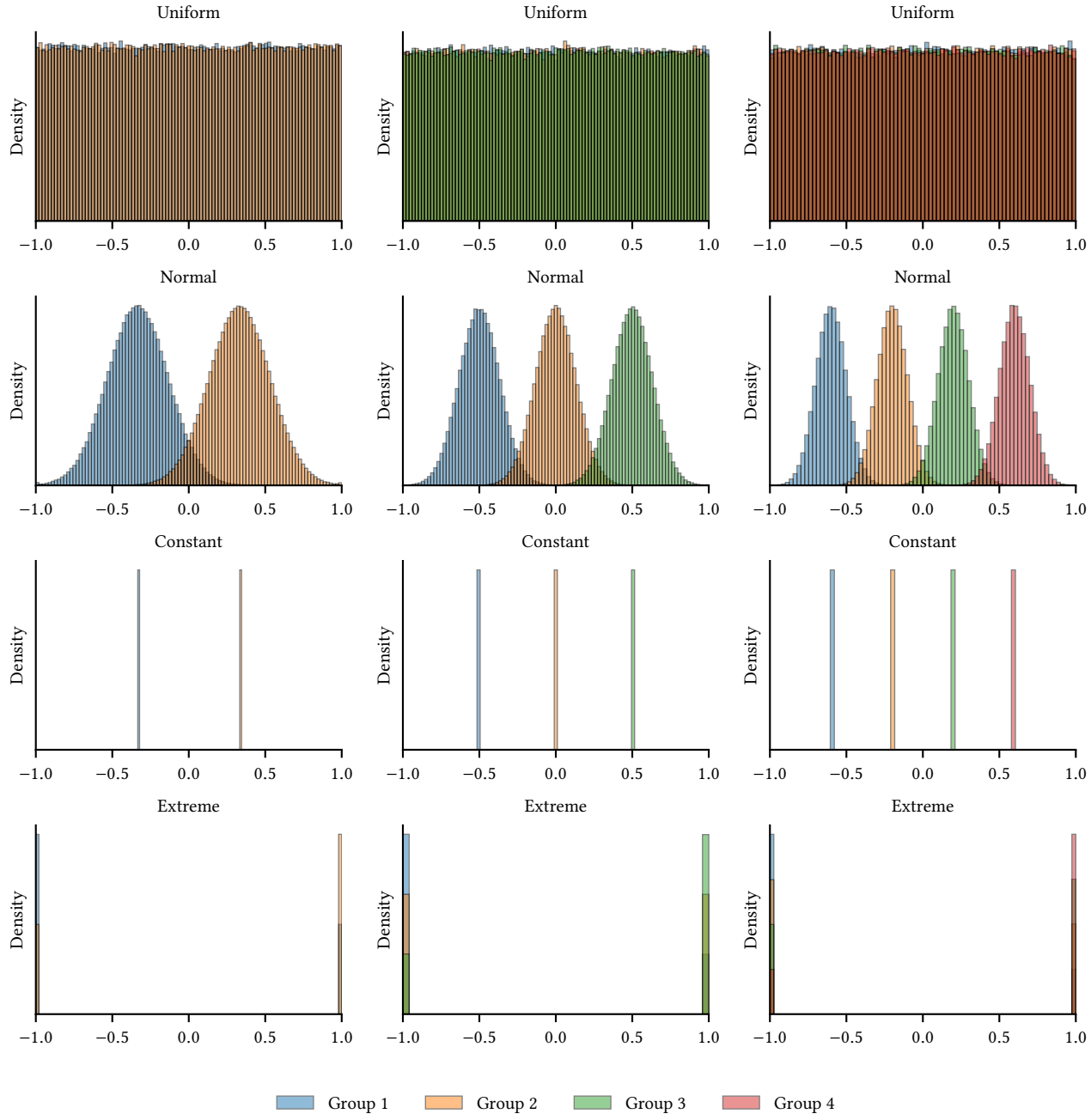


Figure 21: Distribution of the different synthetic datasets used in this paper. First column: two groups. Second column: three groups. Third column: four groups.

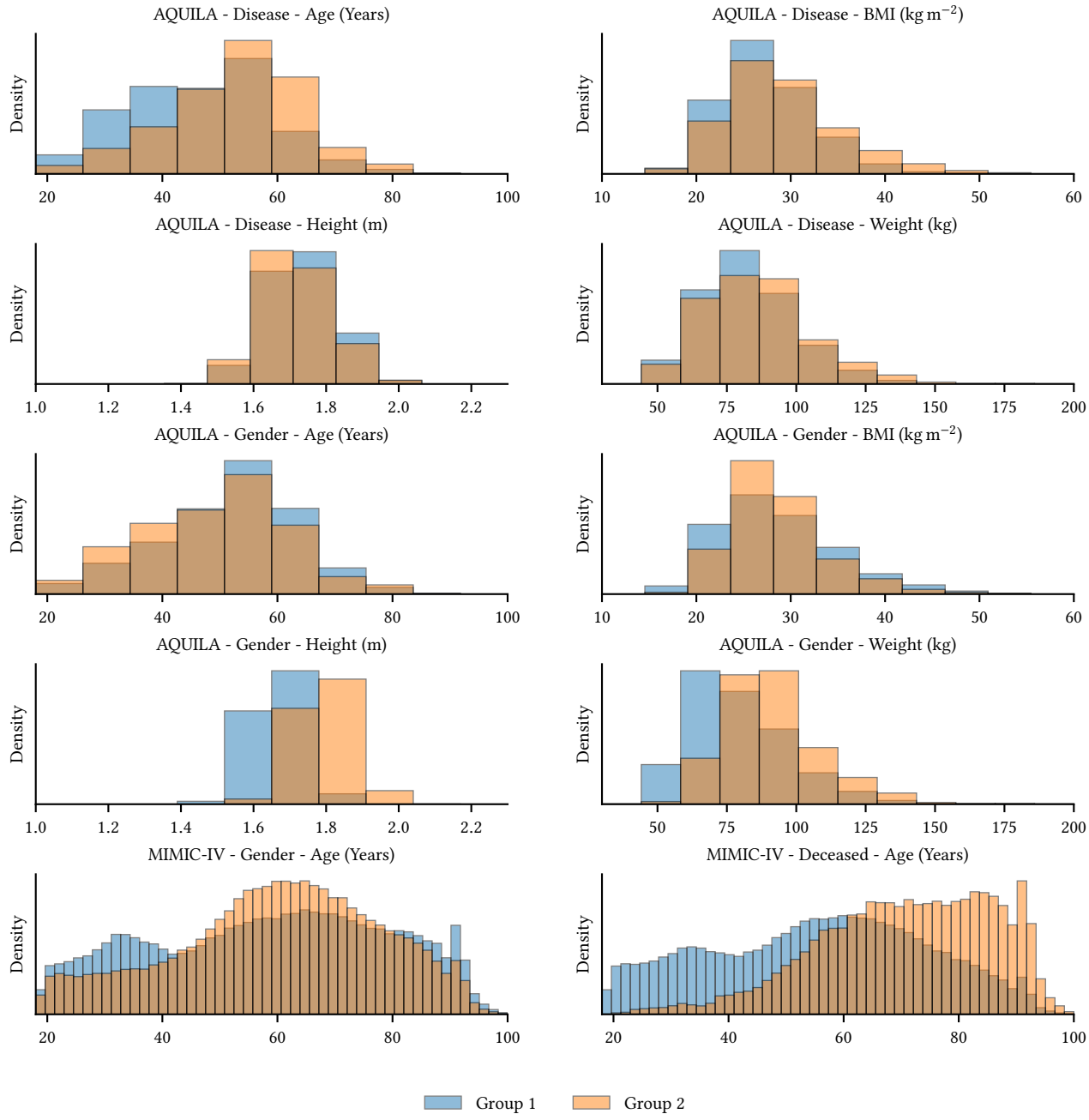


Figure 22: Distribution of the AQUILA and MIMIC-IV datasets used in this paper.

B Notes on the Simplified NPRR Mechanism

The general interactive NPRR mechanism [46] is given by the following algorithm (for fixed $G_t = G$ and $r_t = r$):

Algorithm 4 NPRR mechanism [46]

Input: Value $v \in [0,1]$, parameters $G \in \{1,2,\dots\}$, $r \in (0,1]$
Output: Perturbed value $v' \in \{0, \frac{1}{G}, \frac{2}{G}, \dots, \frac{G}{G}\}$

$v^c \leftarrow \lceil v \cdot G \rceil / G$; $v^f \leftarrow \lfloor v \cdot G \rfloor / G$
if $v^c = v^f$ **then**
  $y \leftarrow v$
else
  $y \sim \begin{cases} v^c & \text{with probability } G \cdot (v - v^f) \\ v^f & \text{with probability } G \cdot (v^c - v) \end{cases}$
end if
$U \sim \text{Uniform}\{0, \frac{1}{G}, \frac{2}{G}, \dots, \frac{G}{G}\}$
$v' \sim \begin{cases} y & \text{with probability } r \\ U & \text{with probability } 1 - r \end{cases}$
return v'

We can simplify this algorithm, by making it non-interactive and making some assumptions that fit the use case in this paper. We assume that ε and G are given and set $r = \frac{e^\varepsilon - 1}{e^\varepsilon + G}$ to ensure that the mechanism is ε-LDP as discussed in [46]. Following this assumption, we can replace the final sampling step of v' with the GRR mechanism for domain size $G+1$ and privacy parameter ε.

Furthermore, we can replace the if-else block with a single sampling operation. We sample a new Bernoulli random variable B with probability $G \cdot v - v^f$ and set $v' = v^f + B/G$. Notice that $B = 0$ corresponds to v^f and $B = 1$ corresponds to v^c. Further notice that $B = 0$ always if $v^c = v^f$ and $y = v$ in this case.

To obtain the simplified algorithm given in Algorithm 2, we apply these changes, rename G to k to avoid confusion, and transform the input range $[0,1]$ to $[-1,1]$ to more closely align with the other algorithms. We perform the transformation by replacing every occurrence of v to $\frac{v+1}{2}$ and the output v' to $2v' - 1$.

C Notes on the data transformation

All algorithms are defined for input in $[-1,1]$, but real data is from $[r,s] \subset \mathbb{R}$. Therefore, we define a function to transform the true input $x \in [r,s]$ to the mechanism input $v = T(x) \in [-1,1]$ and a function to transform the estimated sum $\hat{s}_g$ of transformed values back to the correct value range.

We set $T(x) = 2\frac{x-r}{s-r} - 1$ to transform from $[r,s]$ to $[-1,1]$. We then set $v_i = T(x_i)$ and use v_i as input to a mechanism. We now want to find a function $\bar{T}(y)$ that transforms the resulting sum (over mechanism outputs for the values in $[-1,1]$) back to the correct sum for values in $[r,s]$.

By choosing $\bar{T}(y) = \frac{s-r}{2}y + \hat{n}_g(\frac{s+r}{2})$, we get:

$$\mathbb{E}\left[\bar{T}(\hat{s}_g)\right] = \mathbb{E}\left[\frac{s-r}{2}\hat{s}_g + \hat{n}_g\Big(\frac{s+r}{2}\Big)\right]$$

By the Linearity of Expectation, we get:

$$= \frac{s-r}{2}\mathbb{E}\left[\hat{s}_g\right] + \mathbb{E}\left[\hat{n}_g\right]\left(\frac{s+r}{2}\right)$$

Inserting the expecations for the sum and the count:

$$= \frac{s-r}{2}\Big(\sum_{i=1}^{n_g} v_i\Big) + n_g\Big(\frac{s+r}{2}\Big)$$

Replacing $v_i = T(x_i)$ (see above):

$$= \Big(\sum_{i=1}^{n_g} \frac{s-r}{2}T(x_i)\Big) + n_g\Big(\frac{s+r}{2}\Big)$$

Insert the definition of $T(x)$:

$$= \Big(\sum_{i=1}^{n_g} \frac{s-r}{2}\Big(2\frac{x_i - r}{s-r} - 1\Big)\Big) + n_g\Big(\frac{s+r}{2}\Big)$$

$$= \Big(\sum_{i=1}^{n_g} x_i - r - \frac{s-r}{2}\Big) + n_g\Big(\frac{s+r}{2}\Big)$$

$$= \Big(\sum_{i=1}^{n_g} x_i\Big) - n_g\Big(r + \frac{s-r}{2}\Big) + n_g\Big(\frac{s+r}{2}\Big)$$

$$= \Big(\sum_{i=1}^{n_g} x_i\Big) - n_g\Big(\frac{s+r}{2}\Big) + n_g\Big(\frac{s+r}{2}\Big)$$

$$= \sum_{i=1}^{n_g} x_i$$

Consequently, the application of the two transfer functions to the raw inputs and the unbiased sum estimate leads to an unbiased estimate of the correct sum.

In practice, the use of two estimators ($\hat{s}_g$ AND $\hat{n}_g$) probably leads to additional noise compared to untransformed data in $[-1,1]$ when estimating a sum. Note that for symmetric intervals $[s,r]$, where $s = -r$, the fraction $\frac{s+r}{2}$ becomes 0, and $\hat{n}_g$ does not impact the result at all.

Furthermore, this problem only affects the sum estimation and is not relevant when estimating a mean as $\hat{m}_g = \frac{1}{\hat{n}_g}\bar{T}(\hat{s}_g)$. The expectation of this mean estimate is

$$\frac{1}{\hat{n}_g}\bar{T}(\hat{s}_g) = \frac{1}{\hat{n}_g}\Big(\frac{s-r}{2}\hat{s}_g + \hat{n}_g\Big(\frac{s+r}{2}\Big)\Big)$$

$$= \frac{s-r}{2}\frac{1}{\hat{n}_g}\hat{s}_g + \frac{\hat{n}_g}{\hat{n}_g}\Big(\frac{s+r}{2}\Big)$$

$$= \frac{s-r}{2}\frac{1}{\hat{n}_g}\hat{s}_g + \frac{s+r}{2}$$

We see that we do not have an additive term that is based on the count estimator for the transformed mean estimation. Instead, we only have a constant factor and a constant offset to transform the mean estimate to the correct value range.

D Proofs

D.1 Proof for the group count estimator

Proof of Proposition 1. To show that our estimator $\hat{n}_g$ is an unbiased estimator of the true count n_g or participants in group g, we calculate $\mathbb{E}\left[\hat{n}_g\right]$ and demonstrate that it equals n_g.

For this proof we need the expectation of an indicator function. This expectation is equal to the probability of the event it indicates, i.e., $\mathbb{E}\left[\mathbf{1}_{g'_i=g}\right] = \Pr[g'_i = g]$.

The probability $\Pr[g_i' = g]$ depends on the true group of participant i. That is, if participant i is in group g (i.e., $g_i = g$), they will respond with $g_i' = g$ with probability $a = \frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}$. If their true group is not g (i.e., $g_i \neq g$), they will respond with $g_i' = g$ with probability $q = \frac{1-a}{d-1}$. Therefore, we have:

$$\Pr[g_i' = g] = \begin{cases} a & \text{if } g_i = g, \\ q & \text{if } g_i \neq g, \end{cases}$$

We now calculate the expectation of our estimator:

$$\mathbb{E}\left[\hat{n}_g\right]$$
$$= \mathbb{E}\left[\frac{1}{a-q}\left(\sum_{i=1}^{n} \mathbf{1}_{g_i'=g} - nq\right)\right]$$

Linearity of Expectation:

$$= \frac{1}{a-q}\mathbb{E}\left[\sum_{i=1}^{n} \mathbf{1}_{g_i'=g} - nq\right]$$

Linearity of Expectation:

$$= \frac{1}{a-q}\left(\mathbb{E}\left[\sum_{i=1}^{n} \mathbf{1}_{g_i'=g}\right] - nq\right)$$

Linearity of Expectation:

$$= \frac{1}{a-q}\left(\sum_{i=1}^{n} \mathbb{E}\left[\mathbf{1}_{g_i'=g}\right] - nq\right)$$

Expectation of the indicator function:

$$= \frac{1}{a-q}\left(\sum_{i=1}^{n} \Pr[g_i' = g] - nq\right)$$

Split the sum into the two possible cases ($g_i = g$ and $g_i \neq g$):

$$= \frac{1}{a-q}\left(\left(\sum_{i=1}^{n_g} a + \sum_{j=n_g+1}^{n} q\right) - nq\right)$$
$$= \frac{1}{a-q}\left(n_g a + (n-n_g)q - nq\right)$$
$$= \frac{1}{a-q}\left(n_g a - n_g q\right)$$
$$= \frac{1}{a-q} n_g (a-q)$$
$$= n_g$$

□

D.2 Proofs for the Group-wise Laplace-based Mean Estimation

Proof of Theorem 1. For completeness, we reproduce the proof by Juarez and Korolova [31] for the group-wise Laplace-based mean estimation mechanism and give additional intermediate steps and explanations.

Let $x_0 = (g_0, v_0)$ and $x_1 = (g_1, v_1)$ be two different inputs and $y = (g', v')$ be an output of the mechanism. From the mechanism's definition, we have that for any input $x = (g, v)$,

$$\Pr[y \mid x] = \begin{cases} \frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v'-v) & \text{if } g' = g \\ \frac{1}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v') & \text{if } g' \neq g \end{cases}$$

Here $f_{\mathrm{Lap}(0,\lambda)}(x)$ denotes the probability density function of the Laplace distribution given by $f_{\mathrm{Lap}(0,\lambda)}(x) = \frac{1}{2b} e^{-\frac{|x|}{\lambda}}$ for $x \in \mathbb{R}$ and $\lambda > 0$. The reasoning behind these probabilities is that, when the mechanism preserves the group, $v' = v + Y$ where $Y \sim \mathrm{Lap}(0, \frac{2}{\varepsilon_2})$. Therefore, the probability of observing v' is the probability of sampling $v' - v$ from a Laplace distribution with zero mean and scale $\frac{2}{\varepsilon_2}$. When the mechanism changes the group, it sets $v = 0$ and samples v' from $Lap(0, \frac{2}{\varepsilon_2})$ as well.

If x_0 and x_1 have the same group, we consider two cases: $g' = g$ (Case 1) and $g' \neq g$ (Case 2). If x_0 and x_1 differ in group, we consider two cases: $g' = g_0 \neq g_1$ (Case 3) and $g' = g_1 \neq g_0$ (Case 4).

Case 1: $g' = g$**.** Using the probability of $\Pr[y \mid x]$ when $g' = g$ for both x_0 and x_1, we obtain

$$\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} = \frac{\frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v'-v_0)}{\frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v'-v_1)}$$
$$= \frac{f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v'-v_0)}{f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v'-v_1)}$$
$$= \frac{\frac{1}{4/\varepsilon_2} e^{-\frac{|v'-v_0|}{2/\varepsilon_2}}}{\frac{1}{4/\varepsilon_2} e^{-\frac{|v'-v_1|}{2/\varepsilon_2}}}$$
$$= e^{\frac{|v'-v_1|}{2/\varepsilon_2} - \frac{|v'-v_0|}{2/\varepsilon_2}}$$
$$= e^{\varepsilon_2\left(\frac{|v'-v_1|-|v'-v_0|}{2}\right)}$$

Applying the Triange inequality:

$$\leq e^{\varepsilon_2\left(\frac{|v_0-v_1|}{2}\right)}$$

Using $|v_0|, |v_1| \leq 1$:

$$\leq e^{\frac{2\varepsilon_2}{2}}$$
$$= e^{\varepsilon_2}$$

Case 2: $g' \neq g$**.** Using the probability of $\Pr[y \mid x]$ when $g' \neq g$ for both x_0 and x_1, we obtain

$$\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} = \frac{\frac{1}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v')}{\frac{1}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v')}$$
$$= 1$$
$$= e^0$$

Case 3: $g' = g_0 \neq g_1$**.** Using the probability of $\Pr[y \mid x]$ when $g' = g_0$ for x_0 and $g' \neq g_1$ for x_1, we obtain

$$\begin{aligned}
\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} &= \frac{\frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v'-v_0)}{\frac{1}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v')} \\
&= \frac{e^{\varepsilon_1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v'-v_0)}{f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v')} \\
&= \frac{e^{\varepsilon_1} \frac{1}{4/\varepsilon_2} e^{\frac{-|v'-v_0|}{2/\varepsilon_2}}}{\frac{1}{4/\varepsilon_2} e^{\frac{-|v'|}{2/\varepsilon_2}}} \\
&= e^{\varepsilon_1} e^{\frac{|v'|}{2/\varepsilon_2} - \frac{|v'-v_0|}{2/\varepsilon_2}} \\
&= e^{\varepsilon_1} e^{\varepsilon_2\left(\frac{|v'|}{2} - \frac{|v'-v_0|}{2}\right)} \\
&= e^{\varepsilon_1+\varepsilon_2\left(\frac{|v'|-|v'-v_0|}{2}\right)}
\end{aligned}$$

Applying the Triangle Inequality:

$$\leq e^{\varepsilon_1+\varepsilon_2\left(\frac{|v_0|}{2}\right)}$$

Using $|v_0| \leq 1$:

$$\leq e^{\varepsilon_1+\frac{\varepsilon_2}{2}}$$

Case 4: $g' = g_1 \neq g_0$**.** Using the probability of $\Pr[y \mid x]$ when $g' \neq g_0$ for x_0 and $g' = g_1$ for x_1, we obtain

$$\begin{aligned}
\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} &= \frac{\frac{1}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v')}{\frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v'-v_1)} \\
&= \frac{f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v')}{e^{\varepsilon_1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v'-v_1)} \\
&= \frac{\frac{1}{4/\varepsilon_2} e^{\frac{-|v'|}{2/\varepsilon_2}}}{e^{\varepsilon_1} \frac{1}{4/\varepsilon_2} e^{\frac{-|v'-v_1|}{2/\varepsilon_2}}} \\
&= e^{\frac{|v'-v_1|}{2/\varepsilon_2} - \frac{|v'|}{2/\varepsilon_2} - \varepsilon_1} \\
&= e^{\varepsilon_2 \frac{|v'-v_1|-|v'|}{2} - \varepsilon_1}
\end{aligned}$$

Applying the Triangle Inequality:

$$\leq e^{\varepsilon_2 \frac{|v_1|}{2} - \varepsilon_1}$$

Using $|v_1| \leq 1$:

$$\leq e^{\frac{\varepsilon_2}{2} - \varepsilon_1}$$

Combining all cases, we have that the mechanism is ε-locally differentially private with $\varepsilon = \max\{\varepsilon_2, \varepsilon_1 + \frac{\varepsilon_2}{2}, \frac{\varepsilon_2}{2} - \varepsilon_1\} = \max\{\varepsilon_2, \varepsilon_1 + \frac{\varepsilon_2}{2}\}$. □

Proof of Proposition 2. We show that using different noise scales for the value perturbation when the group is preserved and when the group is changed does not produce a locally differentially private mechanism.

Let $x_0 = (g_0, v_0)$ and $x_1 = (g_1, v_1)$ be two different inputs and $y = (g', v')$ be an output of the mechanism. We keep the default noise scale of $\frac{2}{\varepsilon_2}$ for the value perturbation when the group is preserved and set the noise scale to $\frac{k}{\varepsilon_2}$ when the group is changed. We have that for any input $x = (g, v)$,

$$\Pr[y \mid x] = \begin{cases} \frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v'-v) & \text{if } g' = g \\ \frac{1}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{k}{\varepsilon_2})}(v') & \text{if } g' \neq g \end{cases}$$

For a mechanism to be considered ε-locally differentially private, the ratio of the probabilities of any two outputs y_0 and y_1 given two inputs x_0 and x_1 must be bounded by e^{ε}. We now show that in some cases, there is no bound on this ratio. Consider the case where x_0 and x_1 differ in group and $g' = g_0 \neq g_1$.

We have that

$$\begin{aligned}
\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} &= \frac{\frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v'-v_0)}{\frac{1}{e^{\varepsilon_1}+d-1} f_{\mathrm{Lap}(0,\frac{k}{\varepsilon_2})}(v')} \\
&= \frac{e^{\varepsilon_1} f_{\mathrm{Lap}(0,\frac{2}{\varepsilon_2})}(v'-v_0)}{f_{\mathrm{Lap}(0,\frac{k}{\varepsilon_2})}(v')} \\
&= \frac{e^{\varepsilon_1} \frac{1}{4/\varepsilon_2} e^{\frac{-|v'-v_0|}{2/\varepsilon_2}}}{\frac{1}{2k/\varepsilon_2} e^{\frac{-|v'|}{k/\varepsilon_2}}} \\
&= \frac{e^{\varepsilon_1} \frac{1}{2} e^{\frac{-|v'-v_0|}{2/\varepsilon_2}}}{\frac{1}{k} e^{\frac{-|v'|}{k/\varepsilon_2}}} \\
&= \frac{k}{2} e^{\varepsilon_1} e^{\frac{|v'|}{k/\varepsilon_2} - \frac{|v'-v_0|}{2/\varepsilon_2}} \\
&= \frac{k}{2} e^{\varepsilon_1+\varepsilon_2\left(\frac{2|v'|-k|v'-v_0|}{2k}\right)}
\end{aligned}$$

Applying the Triangle Inequality:

$$\leq \frac{k}{2} e^{\varepsilon_1+\varepsilon_2\left(\frac{|2v'-kv'+kv_0|}{2k}\right)}$$

Using $|v_0| \leq 1, k > 0$:

$$\leq \frac{k}{2} e^{\varepsilon_1+\varepsilon_2\left(\frac{k+|(2-k)v'|}{2k}\right)}$$

Next, we need to find an upper bound for $\frac{k+|(2-k)v'|}{2k}$. Choosing $k = 2$, we have $\frac{k+|v'(2-k)|}{2k} = \frac{k}{2k} = \frac{1}{2}$. In all other cases, this term depends on v' and can be arbitrarily large as $v' \in \mathbb{R}$. Therefore, the mechanism is not locally differentially private for $k \neq 2$. □

Proof of Proposition 3. Our proof for the sum estimate is based on the proof of the mean estimator by Juarez and Korolova [31].

We show that the estimator $\hat{s}_g^{\mathrm{Lap}}$ is unbiased. We model the term $\mathbf{1}_{g'_i=g} \cdot v'_i$ as a random variable V_i:

(1) $g_i = g$: $V_i = G_i(v_i + Y_i)$
(2) $g_i \neq g$: $V_i = \bar{G}_i Y_i$

The random variables $G_i \sim \text{Bernoulli}(a)$ and $\bar{G}_i \sim \text{Bernoulli}(\frac{1-a}{d-1})$ model the group perturbation. The random variables $Y_i \sim \text{Lap}(0, \frac{2}{\varepsilon_2})$ model the Laplace mechanism.

We can now calculate the expectation of the estimator:

$$\mathbb{E}\left[\hat{s}_g^{\text{Lap}}\right] = \mathbb{E}\left[\frac{1}{a}\sum_{i=1}^{n} V_i\right]$$

Linearity of Expectation:

$$= \frac{1}{a}\sum_{i=1}^{n} \mathbb{E}\left[V_i\right]$$

Splitting the sum into the two possible cases:

$$= \frac{1}{a}\left(\sum_{i=1}^{n_g} \mathbb{E}\left[G_i(v_i + Y_i)\right] + \sum_{i=n_g+1}^{n} \mathbb{E}\left[\bar{G}_i Y_i\right]\right)$$

Independence of random variables:

$$= \frac{1}{a}\left(\sum_{i=1}^{n_g} \mathbb{E}\left[G_i\right]\mathbb{E}\left[v_i + Y_i\right] + \sum_{i=n_g+1}^{n} \mathbb{E}\left[\bar{G}_i\right]\mathbb{E}\left[Y_i\right]\right)$$

Linearity of Expecation, $\mathbb{E}\left[G_i\right] = a$ and $\mathbb{E}\left[Y_i\right] = 0$:

$$\begin{aligned}
&= \frac{1}{a}\left(\sum_{i=1}^{n_g} a(v_i + 0) + \sum_{i=n_g+1}^{n} 0\right) \\
&= \frac{1}{a}\sum_{i=1}^{n_g} a v_i \\
&= \sum_{i=1}^{n_g} v_i \\
&= s_g
\end{aligned}$$

□

D.3 Proofs for Group-wise Bernoulli-based Mean Estimation

Proof for Theorem 2. We reproduce the original proof by Juarez and Korolova [31], give some additional explanations and extend it to the general case of d groups. Furthermore, we provide a correction to the original proof that results in a different privacy guarantee.

As in the original proof, we denote $a = \frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}$ and $b = \frac{e^{\varepsilon_2}}{1+e^{\varepsilon_2}}$. Let $x_0 = (g_0, v_0)$ and $x_1 = (g_1, v_1)$ be two different inputs and $y = (g', v')$ be an output of the mechanism.

Let $x_0 = (g_0, v_0)$ and $x_1 = (g_1, v_1)$ be two different inputs and $y = (g', v')$ be an output of the mechanism. From the mechanism's definition, we have that for any input $x = (g, v)$,

$$\Pr[y \mid x] = \begin{cases} \frac{a(1+(2b-1)v'v)}{2} & \text{if } g' = g \\ \frac{1-a}{2(d-1)} & \text{if } g' \neq g \end{cases}$$

Since $v \in [-1, 1]$, and $v' \in \{-1, 1\}$, we can bound the term $v'v$ as $-1 \leq v'v \leq 1$. Using this fact, an upper bound of $\Pr[y \mid x]$ for $g' = g$ is

$$\Pr[y \mid x] \leq \frac{a(1+2b-1)}{2} = ab \tag{10}$$

and a lower bound is

$$\Pr[y \mid x] \geq \frac{a(1-2b+1)}{2} = a(1-b). \tag{11}$$

For the case where the two inputs x_0 and x_1 do not differ in group, the proof is identical to the original proof as d does not appear in the calculation. We reproduce this part of the proof for completeness.

In the following, we bound $\Pr[y \mid x_0]/\Pr[y \mid x_1]$ and distinguish four cases. If x_0 and x_1 have the same group, we consider two cases: $g' = g$ (Case 1) and $g' \neq g$ (Case 2). If x_0 and x_1 differ in group, we consider two cases: $g' = g_0 \neq g_1$ (Case 3) and $g' = g_1 \neq g_0$ (Case 4).

Case 1: $g' = g$**.** Using the upper and lower bounds for $\Pr[y \mid x_0]$ and $\Pr[y \mid x_1]$ respectively, we obtain

$$\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} \leq \frac{ab}{a(1-b)} = \frac{b}{1-b} = e^{\varepsilon_2} \tag{12}$$

Case 2: $g' \neq g$**.** Using the probability of $\Pr[y \mid x]$ when $g' \neq g$ for both x_0 and x_1, we obtain

$$\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} = 1 \leq e^{\varepsilon_2} \tag{13}$$

Case 3: $g' = g_0 \neq g_1$**.** Using the upper bound for $\Pr[y \mid x_0]$, we obtain

$$\begin{aligned}
\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} &= \frac{ab}{\frac{1-a}{2(d-1)}} \\
&\leq \frac{2(d-1)ab}{1-a} \\
&= 2(d-1)\frac{\left(\frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}\right)\left(\frac{e^{\varepsilon_2}}{1+e^{\varepsilon_2}}\right)}{\frac{d-1}{e^{\varepsilon_1}+d-1}} \\
&= 2\frac{e^{\varepsilon_1}e^{\varepsilon_2}}{1+e^{\varepsilon_2}} \\
&= e^{\varepsilon_1 + \ln\left(\frac{2e^{\varepsilon_2}}{e^{\varepsilon_2}+1}\right)}
\end{aligned}$$

Note that this result differs from the original proof by Juarez and Korolova [31] and gives a worse privacy guarantee. The original proof incorrectly states that this ratio can be bounded by e^{ε_1}, but this only holds for $\varepsilon_2 = 0$.

Case 4: $g' = g_1 \neq g_0$**.** Using the lower bound for $\Pr[y \mid x_0]$ and that $1 \leq e^{\varepsilon_2}$, we obtain

$$\begin{aligned}
\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} &\leq \frac{\frac{1-a}{2(d-1)}}{a(1-b)} \\
&= \frac{\frac{1}{2(d-1)}\frac{d-1}{e^{\varepsilon_1}+d-1}}{\frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}\left(1-\frac{e^{\varepsilon_2}}{1+e^{\varepsilon_2}}\right)} \\
&= \frac{\frac{1}{2(d-1)}\frac{d-1}{e^{\varepsilon_1}+d-1}}{\frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}\frac{1}{1+e^{\varepsilon_2}}} \\
&= \frac{1}{2(d-1)}\frac{d-1}{e^{\varepsilon_1}+d-1}\frac{e^{\varepsilon_1}+d-1}{e^{\varepsilon_1}}(1+e^{\varepsilon_2}) \\
&= \frac{1+e^{\varepsilon_2}}{2e^{\varepsilon_1}} \\
&\leq \frac{2e^{\varepsilon_2}}{2e^{\varepsilon_1}} \\
&= e^{\varepsilon_2-\varepsilon_1}
\end{aligned}$$

Again, we see that d cancels out, and we obtain the same result as for $d = 2$ in the original proof.

Combining the upper bounds of the different cases, we can conclude that the mechanism is ε-LDP with $\varepsilon = \max\{\varepsilon_2, \varepsilon_1 + \ln\left(\frac{2e^{\varepsilon_2}}{e^{\varepsilon_2}+1}\right), \varepsilon_2 - \varepsilon_1\} = \max\{\varepsilon_2, \varepsilon_1 + \ln\left(\frac{2e^{\varepsilon_2}}{e^{\varepsilon_2}+1}\right)\}$. □

Proof of Proposition 4. We prove that the estimator $\hat{s}_g^{\text{Bern}}$ is unbiased.

We model the term $\mathbf{1}_{g'=g} \cdot v_i'$ as a random variable V_i:

$$(1)\ g_i = g : V_i = G_i\Big(2\big(R_iB_i + (1-R_i)(1-B_i)\big) - 1\Big)$$
$$(2)\ g_i \neq g : V_i = \bar{G}_i\Big(2\big(R_i\bar{B}_i + (1-R_i)(1-\bar{B}_i)\big) - 1\Big)$$

The random variables $G_i \sim \text{Bernoulli}(a)$ and $\bar{G}_i \sim \text{Bernoulli}(\frac{1-a}{d-1})$ model the group perturbation. The random variables $R_i \sim \text{Bernoulli}\,(b)$ model the perturbation of the Bernoulli sample using randomized response. The random variables $B_i \sim \text{Bernoulli}\left(\frac{1+v_i}{2}\right)$ and $\bar{B}_i \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ model the Bernoulli sampling for the value perturbation.

The expected value of the estimator $\hat{s}_g^{\text{Bern}}$ is

$$\mathbb{E}\left[\hat{s}_g^{\text{Bern}}\right] = \frac{1}{a(2b-1)} \sum_{i=1}^{n} \mathbb{E}\left[\mathbf{1}_{g_i'=g} \cdot v_i'\right]$$

Linearity of Expecation

$$= \frac{1}{a(2b-1)} \left(\sum_{i=1}^{n_g} \mathbb{E}\left[V_i\right] + \sum_{j=n_g+1}^{n} \mathbb{E}\left[V_j\right] \right)$$

Insert definitions of V_i:

$$= \frac{1}{a(2b-1)} \left(\sum_{i=1}^{n_g} \mathbb{E}\left[G_i(2(R_iB_i + (1-R_i)(1-B_i)) - 1)\right] + \sum_{j=n_g+1}^{n} \mathbb{E}\left[\bar{G}_j(2(R_j\bar{B}_j + (1-R_j)(1-\bar{B}_j)) - 1)\right] \right)$$

Independence of random variables, Linearity of Expectation, Expecations of the random variables:

$$\begin{aligned}
&= \frac{1}{a(2b-1)} \left(\sum_{i=1}^{n_g} a\Big(2\Big(b\frac{1+v_i}{2} + (1-b)\frac{1-v_i}{2}\Big) - 1\Big) + \sum_{j=n_g+1}^{n} \frac{1-a}{d-1}\Big(2\Big(b\frac{1}{2} + (1-b)\frac{1}{2}\Big) - 1\Big) \right) \\
&= \frac{1}{a(2b-1)} \left(\sum_{i=1}^{n_g} a\left(2\left(b\frac{1+v_i}{2} + \frac{1-v_i}{2} - b\frac{1-v_i}{2}\right) - 1\right) + \sum_{j=n_g+1}^{n} \frac{1-a}{d-1}\Big(2\frac{1}{2} - 1\Big) \right) \\
&= \frac{1}{a(2b-1)} \left(\sum_{i=1}^{n_g} a\left(b + bv_i + 1 - v_i - b + bv_i - 1\right) + 0 \right) \\
&= \frac{1}{a(2b-1)} \left(\sum_{i=1}^{n_g} a(2b-1)v_i \right) \\
&= \sum_{i=1}^{n_g} v_i \\
&= s_g
\end{aligned}$$

□

D.4 Proofs for Group-wise NPRR-based Mean Estimation

Proof of Theorem 3. We denote $a = \frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}$ and $b = \frac{e^{\varepsilon_2}}{e^{\varepsilon_2}+k}$, where d is the number of groups and k is the parameter of NPRR. We will use $\frac{1-a}{d-1} = \frac{1}{e^{\varepsilon_1}+d-1}$ and $\frac{1-b}{k} = \frac{1}{e^{\varepsilon_2}+k}$ throughout the proof.

Let $x_0 = (g_0, v_0)$ and $x_1 = (g_1, v_1)$ be two different inputs and $y = (g', v')$ be an output of the mechanism. From the mechanism's definition, we have that for any input $x = (g, v)$,

$$\Pr[y \mid x] = \begin{cases} ab & \text{if } g' = g, v' \in \{v_i^c, v_i^f\} \\ a\frac{1-b}{k} & \text{if } g' = g, v' \notin \{v_i^c, v_i^f\} \\ \frac{1-a}{d-1}\frac{1}{k+1} & \text{if } g' \neq g \end{cases}$$

where $v_i^c = \lceil v \cdot k \rceil / k$ and $v_i^f = \lfloor v \cdot k \rfloor / k$.

Since $a, b \in [0, 1]$, an upper bound of $\Pr[y \mid x]$ for $g' = g$ is

$$\Pr[y \mid x] \leq ab \tag{14}$$

and a lower bound is

$$\Pr[y \mid x] \geq a\frac{1-b}{k} \tag{15}$$

In the following, we bound $\Pr[y \mid x_0]/\Pr[y \mid x_1]$ and distinguish four cases. If x_0 and x_1 have the same group, we consider two cases: $g' = g$ (Case 1) and $g' \neq g$ (Case 2). If x_0 and x_1 differ in group, we consider two cases: $g' = g_0 \neq g_1$ (Case 3) and $g' = g_1 \neq g_0$ (Case 4).

Case 1: $g' = g$**.** Using the upper and lower bounds for $\Pr[y \mid x_0]$ and $\Pr[y \mid x_1]$, we obtain

$$\begin{aligned}\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} &\leq \frac{ab}{a\frac{1-b}{k}} \\ &= \frac{\frac{e^{\varepsilon_2}}{e^{\varepsilon_2}+k}}{\frac{1}{e^{\varepsilon_2}+k}} \\ &= e^{\varepsilon_2}\end{aligned}$$

Case 2: $g' \neq g$**.** Using the probability of $\Pr[y \mid x]$ when $g' \neq g$ for both x_0 and x_1, we obtain

$$\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} = 1 \leq e^{\varepsilon_2} \tag{16}$$

Case 3: $g' = g_0 \neq g_1$**.** Using the upper bound for $\Pr[y \mid x_0]$, we obtain

$$\begin{aligned}\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} &\leq \frac{ab}{\frac{1-a}{(k+1)(d-1)}} \\ &= (k+1)\frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}\frac{e^{\varepsilon_2}}{e^{\varepsilon_2}+k}\frac{1}{\frac{1}{e^{\varepsilon_1}+d-1}} \\ &= (k+1)\frac{e^{\varepsilon_1}e^{\varepsilon_2}}{e^{\varepsilon_2}+k}\end{aligned}$$

One possible upper bound for this term is $e^{\varepsilon_1+\varepsilon_2}$. However, we can get a tighter bound, by keeping the term as

$$\begin{aligned}(k+1)\frac{e^{\varepsilon_1}e^{\varepsilon_2}}{e^{\varepsilon_2}+k} &= e^{\varepsilon_1}\frac{(k+1)e^{\varepsilon_2}}{e^{\varepsilon_2}+k} \\ &= e^{\varepsilon_1+\ln\left(\frac{(k+1)e^{\varepsilon_2}}{e^{\varepsilon_2}+k}\right)}\end{aligned}$$

Therefore, we can bound the ratio as

$$\begin{aligned}\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} &\leq \min\left\{e^{\varepsilon_1+\varepsilon_2}, e^{\varepsilon_1+\ln\left(\frac{(k+1)e^{\varepsilon_2}}{e^{\varepsilon_2}+k}\right)}\right\} \\ &= e^{\varepsilon_1+\ln\left(\frac{(k+1)e^{\varepsilon_2}}{e^{\varepsilon_2}+k}\right)}\end{aligned} \tag{17}$$

Case 4: $g' = g_1 \neq g_0$**.** Using the lower bound for $\Pr[y \mid x_0]$, we obtain

$$\begin{aligned}\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} &\leq \frac{\frac{1-a}{(k+1)(d-1)}}{a\frac{1-b}{k}} \\ &= \frac{k}{(k+1)(d-1)}\frac{1-a}{a(1-b)} \\ &= \frac{k}{(k+1)(d-1)}\frac{\frac{d-1}{e^{\varepsilon_1}+d-1}}{\frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}\frac{k}{e^{\varepsilon_2}+k}} \\ &= \frac{k}{(k+1)(d-1)}\frac{d-1}{e^{\varepsilon_1}+d-1}\frac{e^{\varepsilon_1}+d-1}{e^{\varepsilon_1}}\frac{e^{\varepsilon_2}+k}{k} \\ &= \frac{e^{\varepsilon_2}+k}{(k+1)e^{\varepsilon_1}} \\ &\leq \frac{e^{\varepsilon_2}+ke^{\varepsilon_2}}{(k+1)e^{\varepsilon_1}} \qquad \text{(Using } 1 \leq e^{\varepsilon_2}\text{)} \\ &= \frac{(k+1)e^{\varepsilon_2}}{(k+1)e^{\varepsilon_1}} \\ &= e^{\varepsilon_2-\varepsilon_1}\end{aligned}$$

Combining the upper bounds of the different cases, we can conclude that the mechanism is ε-LDP with

$$\begin{aligned}\varepsilon &= \max\left\{\varepsilon_2, \varepsilon_1 + \ln\left(\frac{(k+1)e^{\varepsilon_2}}{e^{\varepsilon_2}+k}\right), \varepsilon_2 - \varepsilon_1\right\} \\ &= \max\left\{\varepsilon_2, \varepsilon_1 + \ln\left(\frac{(k+1)e^{\varepsilon_2}}{e^{\varepsilon_2}+k}\right)\right\}\end{aligned}$$

□

Proof of Proposition 5. We prove that the estimator $\hat{s}_g^{\text{NPRR}}$ is unbiased.

We model the term $\mathbf{1}_{g'=g} \cdot v_i'$ as a random variable V_i:

(1) $g_i = g : V_i = G_i\Big(2(B_iY_i + (1-B_i)U_i) - 1\Big)$

(2) $g_i \neq g : V_i = \bar{G}_i\Big(2(B_iY_i^0 + (1-B_i)U_i) - 1\Big)$

where G_i and $\bar{G}_j$ model the group GRR as Bernoulli random variables with parameters a and $\frac{1-a}{d-1}$, respectively. The Bernoulli random variables B_i model the value perturbation GRR with parameter $b = \frac{e^{\varepsilon_2}-1}{e^{\varepsilon_2}+k}$. Y_i and Y_i^0 model the private values after the stochastic rounding process for the true value v and the neutral value 0, respectively. The uniform random variables U_i model the uniform sample from all values (including the "correct" one), i.e. $U_i \sim \text{Uniform}(\{0, \frac{1}{k}, \frac{2}{k} \ldots, \frac{k}{k} = 1\})$.

The expected values of these intermediate random variables are

- $\mathbb{E}[G_i] = a$
- $\mathbb{E}\left[\bar{G}_j\right] = \frac{1-a}{d-1}$
- $\mathbb{E}[B_i] = \frac{e^{\varepsilon_2}-1}{e^{\varepsilon_2}+k} = b$
- $Y_i = v_i^f + Y_i^B/k$ where $Y_i^B \sim \text{Bernoulli}(k(\frac{v_i+1}{2} - v_i^f))$. Then $\mathbb{E}[Y_i] = v_i^f + k(\frac{v_i+1}{2} - v_i^f)/k = \frac{v_i+1}{2}$
- $\mathbb{E}\left[Y_i^0\right] = 0$ (from the definition of $\mathbb{E}[Y_i]$ with $v_i = \frac{1}{2}$)
- $\mathbb{E}[U_i] = \frac{0+1}{2} = \frac{1}{2}$

Using this model and the expected values of the intermediate random variables, we get

$$\mathbb{E}\left[\hat{s}_g^{\text{NPRR}}\right] = \mathbb{E}\left[\frac{1}{ab}\Big(\sum_{i=1}^{n}\mathbf{1}_{g_i'=g} \cdot v_i'\Big)\right]$$

Linearity of expectation

$$= \frac{1}{ab}\sum_{i=1}^{n}\mathbb{E}\left[\mathbf{1}_{g_i'=g} \cdot v_i'\right]$$

Split the sum into the two possible cases regarding g_i

$$= \frac{1}{ab}\left(\sum_{i=1}^{n_g}\mathbb{E}[V_i] + \sum_{j=n_g+1}^{n}\mathbb{E}\left[V_j\right]\right)$$

Insert the definitions of V_i for both cases

$$= \frac{1}{ab}\left(\sum_{i=1}^{n_g}\mathbb{E}\left[G_i\Big(2(B_iY_i + (1-B_i)U_i) - 1\Big)\right] + \sum_{j=n_g+1}^{n}\mathbb{E}\left[\bar{G}_i\Big(2(B_iY_i^0 + (1-B_i)U_i) - 1\Big)\right]\right)$$

All remaining random variables are mutually independent

$$= \frac{1}{ab}\left(\sum_{i=1}^{n_g}\mathbb{E}[G_i]\left(2(\mathbb{E}[B_i]\,\mathbb{E}[Y_i]+(1-\mathbb{E}[B_i])\mathbb{E}[U_i])-1\right)+\sum_{j=n_g+1}^{n}\mathbb{E}\left[\bar{G}_j\right]\left(2(\mathbb{E}\left[B_j\right]\mathbb{E}\left[Y_j^0\right]+(1-\mathbb{E}\left[B_j\right])\mathbb{E}\left[U_j\right])-1\right)\right)$$

Insert the expected values of the random variables

$$= \frac{1}{ab}\left(\sum_{i=1}^{n_g} a\left(2\left(b\frac{v_i+1}{2}+(1-b)\cdot\frac{1}{2}\right)-1\right)+\sum_{j=n_g+1}^{n}\frac{1-a}{d-1}\left(2\left(b\cdot\frac{1}{2}+(1-b)\cdot\frac{1}{2}\right)-1\right)\right)$$

$$= \frac{1}{ab}\left(\sum_{i=1}^{n_g} a(bv_i+b+1-b-1)+\sum_{j=n_g+1}^{n}\frac{1-a}{d-1}(b+1-b-1)\right)$$

$$= \frac{1}{ab}\sum_{i=1}^{n_g} abv_i + 0$$

$$= \sum_{i=1}^{n_g} v_i$$

$$= s_g$$

□

D.5 Proofs for Group-wise Piecewise Mean Estimation

Proof of Proposition 6. We proof the privacy of the Group Piecewise mechanism. We denote $a = \frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}$ and $p = \frac{e^{\varepsilon_2}-e^{\varepsilon_2/2}}{2e^{\varepsilon_2/2}+2}$. From the mechanism's definition, we have $C = \frac{e^{\varepsilon_2/2}+1}{e^{\varepsilon_2/2}-1}$, $l(v) = \frac{C+1}{2}\cdot v - \frac{C-1}{2}$, and $r(v) = l(v) + C - 1$.

Let $x_0 = (g_0, v_0)$ and $x_1 = (g_1, v_1)$ be two different inputs and $y = (g', v')$ be an output of the mechanism. From the mechanism's definition, we have that for any input $x = (g, v)$,

$$\Pr[y \mid x] = \begin{cases} ap & \text{if } g' = g, v' \in [l(v), r(v)] \\ a\frac{p}{e^{\varepsilon_2}} & \text{if } g' = g, v' \in [-C, l(v)) \cup (r(v), C] \\ \frac{1-a}{d-1}p & \text{if } g' \neq g, v' \in [-\frac{C-1}{2}, \frac{C-1}{2}] \\ \frac{1-a}{d-1}\frac{p}{e^{\varepsilon_2}} & \text{if } g' \neq g, v' \in [-C, -\frac{C-1}{2}) \cup (\frac{C-1}{2}, C] \end{cases}$$

In the following, we bound $\Pr[y \mid x_0]/\Pr[y \mid x_1]$ and distinguish four cases. If x_0 and x_1 have the same group, we consider two cases: $g' = g$ (Case 1) and $g' \neq g$ (Case 2). If x_0 and x_1 differ in group, we consider two cases: $g' = g_0 \neq g_1$ (Case 3) and $g' = g_1 \neq g_0$ (Case 4).

Case 1: $g' = g_0 = g_1$.

$$\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} \leq \frac{ap}{a\frac{p}{e^{\varepsilon_2}}} = e^{\varepsilon_2}$$

Case 2: $g' \neq g_0 = g_1$.

$$\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} \leq \frac{\frac{1-a}{d-1}p}{\frac{1-a}{d-1}\frac{p}{e^{\varepsilon_2}}} = e^{\varepsilon_2}$$

Case 3: $g' = g_0 \neq g_1$.

$$\begin{aligned}\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} &\leq \frac{ap}{\frac{1-a}{d-1}\frac{p}{e^{\varepsilon_2}}} \\ &= (d-1)\frac{a}{1-a}e^{\varepsilon_2} \\ &= (d-1)\frac{e^{\varepsilon_1}}{e^{\varepsilon_1}+d-1}\frac{e^{\varepsilon_1}+d-1}{d-1}e^{\varepsilon_2} \\ &= e^{\varepsilon_1+\varepsilon_2}\end{aligned}$$

Case 4: $g' = g_1 \neq g_0$.

$$\begin{aligned}\frac{\Pr[y \mid x_0]}{\Pr[y \mid x_1]} &\leq \frac{\frac{1-a}{d-1}p}{a\frac{p}{e^{\varepsilon_2}}} \\ &= \frac{1-a}{(d-1)a}e^{\varepsilon_2} \\ &= \frac{1}{d-1}\frac{d-1}{e^{\varepsilon_1}+d-1}\frac{e^{\varepsilon_1}+d-1}{e^{\varepsilon_1}}e^{\varepsilon_2} \\ &= \frac{e^{\varepsilon_2}}{e^{\varepsilon_1}} \\ &= e^{\varepsilon_2-\varepsilon_1}\end{aligned}$$

Combining the upper bounds of the different cases, we can conclude that the mechanism is ε-LDP with $\varepsilon = \max\{\varepsilon_2, \varepsilon_1+\varepsilon_2, \varepsilon_2-\varepsilon_1\} = \varepsilon_1 + \varepsilon_2$.

□

Proof of Proposition 7. We prove that the estimator $\hat{s}_g^{\mathrm{PW}}$ is unbiased.

We model the term $\mathbf{1}_{g_i'=g}(v_i')$ as a random variable V_i:

(1) $g_i = g$: $V_i = G_i Y_i$
(2) $g_i \neq g$: $V_i = \hat{G}_i \hat{Y}_i$

The random variables $G_i \sim \mathrm{Bernoulli}(a)$ and $\hat{G}_i \sim \mathrm{Bernoulli}(\frac{1-a}{d-1})$ model the group perturbation. The random variables $Y_i = M_{\mathrm{PW}}(v_i, \varepsilon_2)$ and $\hat{Y}_i = M_{\mathrm{PW}}(0, \varepsilon_2)$ model the value perturbation.

We know from the original paper [42], which introduces the Piecewise mechanism, that each mechanism output v' is unbiased, i.e. $\mathbb{E}[Y_i] = v_i$. We will also use the fact that in the case of $g_i \neq g$, we set $v_i = 0$ and therefore $\mathbb{E}\left[\hat{Y}_i\right] = 0$.

$$\mathbb{E}\left[\hat{s}_g^{\mathrm{PW}}\right] = \mathbb{E}\left[\frac{1}{a}\sum_{i=1}^{n}\mathbf{1}_{g_i'=g}\cdot v_i'\right]$$

Linearity of Expectation:

$$= \frac{1}{a}\sum_{i=1}^{n_g}\mathbb{E}[V_i]$$

Splitting the sum into the two possible cases ($g_i = g$ and $g_i \neq g$):

$$= \frac{1}{a}\left(\sum_{i=1}^{n_g}\mathbb{E}[V_i] + \sum_{i=n_g+1}^{n}\mathbb{E}[V_i]\right)$$

$$= \frac{1}{a}\left(\sum_{i=1}^{n_g}\mathbb{E}[G_i Y_i] + \sum_{i=n_g+1}^{n}\mathbb{E}\left[\hat{G}_i\hat{Y}_i\right]\right)$$

Independence of the perturbed values and the group perturbation:

$$= \frac{1}{a}\left(\sum_{i=1}^{n_g} \mathbb{E}\left[G_i\right]\mathbb{E}\left[V_i\right] + \sum_{i=n_g+1}^{n} \mathbb{E}\left[\hat{G}_i\right]\mathbb{E}\left[\hat{Y}_i\right]\right)$$

Expectation of the indicator function + Unbiasedness of the mechanism:

$$\begin{aligned} &= \frac{1}{a}\left(\sum_{i=1}^{n_g} a v_i + \sum_{i=n_g+1}^{n} \frac{1-a}{d-1}\cdot 0\right) \\ &= \sum_{i=1}^{n_g} v_i \\ &= s_g \end{aligned}$$

Therefore, the estimator $\hat{s}_g^{\mathrm{PW}}$ is unbiased. □