

Bias in allocations using differentially private census data: An analysis of the 2020 U.S. Census

Author One^{a,c,1}, Author Two^{b,1,2}, and Author Three^a

This manuscript was compiled on June 7, 2023

The decennial census is a primary source of data for the US government to make critical decisions. For example, 132 programs used Census Bureau data to distribute more than \$675 billion in funds during fiscal year 2015. In order to ensure the published census data does not reveal individual information, the Census Bureau adopts the differential privacy (DP) technique. In particular, in 2020, the Census Bureau implemented the Top Down algorithm which release statistical data from top (nation) to bottom hierarchical level (block). However, it was recently observed that the DP outcomes can introduce biased outcomes, especially for minority groups. In this paper, we analyze the reasons for these disproportionate impacts and proposes guidelines to mitigate these effects. We focus on two aspects that can produce the fair outcomes: (1) shape of allocation function and (2) impact of post-processing steps.

Differential Privacy | Fairness | Resource Allocation | Census

Introduction

Agencies, such as the U.S. Census Bureau, release data sets and statistics about groups of individuals that are then used as inputs to a number critical decision processes. For example, the census data is used to decide whether a jurisdiction must provide language assistance during elections, Title I fund allocation in education (1) and to establish national level COVID-19 vaccination distribution plans for states and jurisdictions (2). The resulting decisions can have significant societal, economic, and medical consequences for participating individuals.

In many cases, the released data contain sensitive information and their privacy is strictly regulated. For example, in the U.S., the census data is regulated under Title 13 (3), which requires that no individual be identified from any data release by the Census Bureau. In Europe, data release are regulated according to the *General Data Protection Regulation* (4), which addresses the control and transfer of personal data.

Statistical agencies thus release *privacy-preserving* data and statistics that conform to privacy and confidentiality requirements. In the U.S., a small number of decisions, such as congressional apportionment, are taken using unprotected true values, but the vast majority of decisions rely on privacy-preserving data. Of particular interest are resource allocation decisions relying on the U.S. Census Bureau data, since the bureau will release several privacy-preserving data products using the framework of *Differential Privacy* (5) for their 2020 release.

In particular, in 2020 the Census Bureau implemented a new privacy preserving framework to release privately hierarchical statistical data, the Top Down algorithm. The algorithms works by firstly splitting the given privacy budget ϵ to six hierarchical levels (nation, state, county, tract, block, group—block). Then in the second step, a post-processing step is applied to make sure the noisy counts are consistent, e.g., the counts should be non-negative. However, (1) empirically showed that differential privacy may have a disparate impact on several resource allocation problems. The noise introduced by the privacy mechanism may result in decisions that impact various groups differently. Unfortunately, the paper did not provide a deep understanding why this behavior happens and any mitigation to resolve the issue. This paper builds on these observations and provides a step towards a deeper understanding of the fairness issues arising when differentially private data is used as input to several resource allocation problems. *One of its main results is to prove that several allotment problems and decision rules with significant societal impact (e.g., the allocation of educational funds, the decision to provide minority language assistance on election ballots, or the distribution of COVID-19 vaccines) exhibit inherent unfairness when applied to a differentially private release of the census data.* To counteract this negative results, the paper examines the conditions under which decision making is fair when using differential privacy, and techniques

Significance Statement

This research addresses a critical gap in understanding the relationship between privacy-preserving data release methods, particularly differential privacy, and the fairness of consequential societal decisions derived from such data. By investigating the impacts of differential privacy on resource allocation, this study unveils the inherent potential for bias against various groups, which could detrimentally affect areas such as educational funding, electoral assistance, and public health initiatives. In response to these findings, our work proposes guidelines to bound unfairness and foster equity, providing a roadmap for future applications of differential privacy in data-driven decision making. The insights gleaned from this study have implications not only for the application of privacy-preserving techniques, but also for broader conversations around data privacy, fairness, and social equity

Author affiliations: ^aAffiliation One; ^bAffiliation Two; ^cAffiliation Three

Please provide details of author contributions here.

Please declare any competing interests here.

¹ A.O.(Author One) contributed equally to this work with A.T. (Author Two) (remove if not applicable).

² To whom correspondence should be addressed. E-mail: author.two@email.com

to bound unfairness. The paper also provides a number of mitigation approaches to alleviate biases introduced by differential privacy on such decision making problems. More specifically, we make the following contributions:

1. We formally defines notions of fairness and bounded fairness for decision making subject to privacy requirements.
2. We examine the roots of the induced unfairness by analyzing the structure of the decision making problems.
3. We propose several guidelines to mitigate the negative fairness effects of the decision problems studied.

To the best of the authors' knowledge, this is the first study that attempt at characterizing the relation between differential privacy and fairness in decision problems. All proofs are reported in the appendix. [Is this still true?]

Background: Differential Privacy

Differential Privacy (6) (DP) is a rigorous privacy notion that characterizes the amount of information of an individual's data being disclosed in a computation.

Definition 1 A randomized algorithm $\mathcal{M} : \mathcal{X} \rightarrow \mathcal{R}$ with domain \mathcal{X} and range \mathcal{R} satisfies ϵ -differential privacy if for any output $O \subseteq \mathcal{R}$ and data sets $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$ differing by at most one entry (written $\mathbf{x} \sim \mathbf{x}'$)

$$\Pr[\mathcal{M}(\mathbf{x}) \in O] \leq \exp(\epsilon) \Pr[\mathcal{M}(\mathbf{x}') \in O]. \quad [1]$$

Parameter $\epsilon > 0$ is the *privacy loss*, with values close to 0 denoting strong privacy. Intuitively, DP states that any event occur with similar probability regardless of the participation of any individual data to the data set.

DP satisfies several properties including *immunity to post-processing*, which states that the privacy loss of DP outputs is not affected by arbitrary data-independent post-processing (7).

A function f from a data set $\mathbf{x} \in \mathcal{X}$ to a result set $R \subseteq \mathbb{R}^n$ can be made differentially private by injecting random noise onto its output. The amount of noise relies on the notion of *global sensitivity* $\Delta_f = \max_{\mathbf{x} \sim \mathbf{x}'} \|f(\mathbf{x}) - f(\mathbf{x}')\|_1$. The *Laplace mechanism* (6) that outputs $f(\mathbf{x}) + \boldsymbol{\eta}$, where $\boldsymbol{\eta} \in \mathbb{R}^n$ is drawn from the i.i.d. Laplace distribution with 0 mean and scale Δ_f/ϵ over n dimensions, achieves ϵ -DP. Similarly, once can also utilize the Gaussian mechanism to satisfy ϵ -DP by sampling from the Gaussian distribution with mean 0 and scale Δ_f/ϵ .

Problem Setting and Goals

The paper considers a dataset $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^k$ of n entities, whose elements $x_i = (x_{i1}, \dots, x_{ik})$ describe k measurable quantities of entity $i \in [n]$, such as the number of individuals living in a geographical region i and their English proficiency. The paper considers two classes of problems:

- An *allotment problem* $P : \mathcal{X} \times [n] \rightarrow \mathbb{R}$ is a function that distributes a finite set of resources to some problem entity. P may represent, for instance, the amount of money allotted to a school district.
- A *decision rule* $P : \mathcal{X} \times [n] \rightarrow \{0, 1\}$ determines whether some entity qualifies for some benefits. For instance, P may represent if election ballots should be described in a minority language for an electoral district.



Fig. 1. Diagram of the private allocation problem.

The paper assumes that P has bounded range, and uses the shorthand $P_i(\mathbf{x})$ to denote $P(\mathbf{x}, i)$ for entity i . The focus of the paper is to study the effects of a DP data-release mechanism \mathcal{M} to the outcomes of problem P . Mechanism \mathcal{M} is applied to the dataset \mathbf{x} to produce a privacy-preserving counterpart $\tilde{\mathbf{x}}$ and the resulting private outcome $P_i(\tilde{\mathbf{x}})$ is used to make some allocation decisions.

Figure 2 provides an illustrative diagram.

Because random noise is added to the original dataset \mathbf{x} , the output $P_i(\tilde{\mathbf{x}})$ incurs some error. *The focus of this paper is to characterize and quantify the disparate impact of this error among the problem entities.* In particular, the paper focuses on two notations of errors.

Definition 2 (Statistical bias) The statistical bias $B_P^i(\mathcal{M}, \mathbf{x})$ of the mechanism \mathcal{M} measures the difference between the expected private outcome with the true outcome:

$$B_P^i(\mathcal{M}, \mathbf{x}) = \mathbb{E}_{\tilde{\mathbf{x}} \sim \mathcal{M}(\mathbf{x})} [P_i(\tilde{\mathbf{x}})] - P_i(\mathbf{x}), \quad [2]$$

The paper also considers another notation of error which is the normalized version of the above bias.

Definition 3 (Multiplicative error) The multiplicative error under mechanism \mathcal{M} and problem P for entity i is given by: $B_P^i(\mathcal{M}, \mathbf{x})/P^i(\mathbf{x})$

Our notion of fairness will be based on these two notations of errors.

Definition 4 (α -fairness (8)) Given the true data \mathbf{x} , the mechanism \mathcal{M} is said to be α -fair if, for any $i \in [n]$,

$$\xi_P^i(\mathcal{M}, \mathbf{x}) = |B_P^i(\mathcal{M}, \mathbf{x}) - B_P^j(\mathcal{M}, \mathbf{x})| \leq \alpha,$$

where $\xi_P^i(\mathcal{M}, \mathbf{x})$ is referred to as the disparity error associated with district i . The mechanism \mathcal{M} is α' -minimally fair if $\alpha' = \inf \alpha$ such that \mathcal{M} is α -fair. To put it differently, the mechanism \mathcal{M} is α' -minimally fair if

$$\begin{aligned} \alpha' &= \max_{j \neq i} |B_P^i(\mathcal{M}, \mathbf{x}) - B_P^j(\mathcal{M}, \mathbf{x})| \\ &= \max_{j \in [n]} B_P^j(\mathcal{M}, \mathbf{x}) - \min_{j \in [n]} B_P^j(\mathcal{M}, \mathbf{x}). \end{aligned}$$

Throughout this report, every time we say that a mechanism is α -fair, we mean that it is α -minimally fair.

Problem Setting and Goals

The paper considers a dataset $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^k$ of n entities, whose elements $x_i = (x_{i1}, \dots, x_{ik})$ describe k measurable quantities of entity $i \in [n]$, such as the number of individuals living in a geographical region i and their English proficiency. The paper considers two classes of problems:

- An *allotment problem* $P : \mathcal{X} \times [n] \rightarrow \mathbb{R}$ is a function that distributes a finite set of resources to some problem entity. P may represent, for instance, the amount of money allotted to a school district.

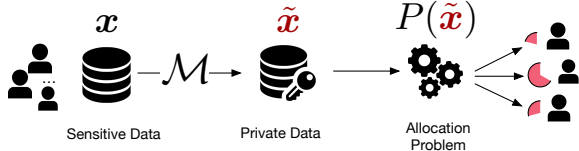


Fig. 2. Diagram of the private allocation problem.

- A decision rule $P : \mathcal{X} \times [n] \rightarrow \{0, 1\}$ determines whether some entity qualifies for some benefits. For instance, P may represent if election ballots should be described in a minority language for an electoral district.

The paper assumes that P has bounded range, and uses the shorthand $P_i(\mathbf{x})$ to denote $P(\mathbf{x}, i)$ for entity i . The focus of the paper is to study the effects of a DP data-release mechanism \mathcal{M} to the outcomes of problem P . Mechanism \mathcal{M} is applied to the dataset \mathbf{x} to produce a privacy-preserving counterpart $\tilde{\mathbf{x}}$ and the resulting private outcome $P_i(\tilde{\mathbf{x}})$ is used to make some allocation decisions.

Figure 2 provides an illustrative diagram.

Because random noise is added to the original dataset \mathbf{x} , the output $P_i(\tilde{\mathbf{x}})$ incurs some error. *The focus of this paper is to characterize and quantify the disparate impact of this error among the problem entities.* In particular, the paper focuses on two notations of errors.

Definition 5 (Statistical bias) *The statistical bias $B_P^i(\mathcal{M}, \mathbf{x})$ of the mechanism \mathcal{M} measures the difference between the expected private outcome with the true outcome:*

$$B_P^i(\mathcal{M}, \mathbf{x}) = \mathbb{E}_{\tilde{\mathbf{x}} \sim \mathcal{M}(\mathbf{x})} [P_i(\tilde{\mathbf{x}})] - P_i(\mathbf{x}), \quad [3]$$

The paper also considers another notation of error which is the normalized version of the above bias.

Definition 6 (Multiplicative error) *The multiplicative error under mechanism \mathcal{M} and problem P for entity i is given by: $B_P^i(\mathcal{M}, \mathbf{x}) / P_i(\mathbf{x})$*

Our notion of fairness will be based on these two notations of errors.

Definition 7 (α -fairness (8)) *Given the true data \mathbf{x} , the mechanism \mathcal{M} is said to be α -fair if, for any $i \in [n]$,*

$$\xi_P^i(\mathcal{M}, \mathbf{x}) = |B_P^i(\mathcal{M}, \mathbf{x}) - B_P^j(\mathcal{M}, \mathbf{x})| \leq \alpha,$$

where $\xi_P^i(\mathcal{M}, \mathbf{x})$ is referred to as the disparity error associated with district i . The mechanism \mathcal{M} is α' -minimally fair if $\alpha' = \inf \alpha$ such that \mathcal{M} is α -fair. To put it differently, the mechanism \mathcal{M} is α' -minimally fair if

$$\begin{aligned} \alpha' &= \max_{j \neq i} |B_P^i(\mathcal{M}, \mathbf{x}) - B_P^j(\mathcal{M}, \mathbf{x})| \\ &= \max_{j \in [n]} B_P^j(\mathcal{M}, \mathbf{x}) - \min_{j \in [n]} B_P^j(\mathcal{M}, \mathbf{x}). \end{aligned}$$

Throughout this report, every time we say that a mechanism is α -fair, we mean that it is α -minimally fair.

Title I School Funding Allocation

A. Allocation Problem. The Title I of the Elementary and Secondary Education Act of 1965 (9) distributes funds through its basic, concentration, and target grants which account for

Table 1. Description of the thresholds that construct the eligibility criteria for each grant.

Grant Type	Eligibility Criteria	Thresholds
Basic Grant	Population of eligible students	> 10
	Proportion of eligible students	> 0.02
Concentration Grant	District population	> 6500
	Proportion of eligible students	> 0.15
Targeted Grant	Population of eligible students	> 10
	Proportion of eligible students	> 0.05

\$6.2B, \$1.3B, and \$4.2B of allocations, respectively. The federal allotment is divided among nearly 17000 qualifying school districts in proportion to the count x_i of children aged 5 to 17 who live in necessitous families in district i . Each grant is specified by a set of thresholds that determines which districts are eligible for that particular grant. We refer to (9) which discusses the specifics of each grant, which we describe briefly here.

The basic allocation grant is formalized by:

$$P_i^F(\mathbf{x}) \stackrel{\text{def}}{=} \left(\frac{x_i \cdot a_i}{\sum_{i \in [n]} x_i \cdot a_i} \right) \cdot F_B \quad [4]$$

where $\mathbf{x} = (x_i)_{i \in [n]}$ is the vector of all eligible districts counts and a_i is a weight factor reflecting students expenditures (defined as the adjusted state per-pupil expenditure). F_B is the total basic Title I appropriation for the fiscal year. Finally, each funding grant is subject to different thresholds which define eligibility criteria for each district. This criteria is described in Table 1. The targeted grant is further weighted by population, as outlined in Section A of (9).

B. Privacy Budget. The Disclosure Avoidance System for the US Census Bureau defines a global rho that is distributed hierarchically. The global privacy-loss budget for 2020 was defined as $\rho = 2.56$. The DAS suggests the use of this bound to convert the ρ -based privacy-loss budgets to the (ϵ, δ) equivalents. $\epsilon = \rho + 2 \cdot \sqrt{-\rho \cdot \log_e \delta}$ where we define $\delta = 10e - 10$. In a 1-year estimate of the ACS, there are 1426 different "Detailed tables" available from the US Census Bureau. The rho allocation for the national, state, and county level geographic hierarchies are $\frac{104}{4099}$, $\frac{1440}{4099}$, and $\frac{447}{4099}$ respectively. Assuming that the county level privacy-loss budget is split equally across these tables, each field would have $\rho = 2.56 \cdot \frac{447}{4099} \cdot \frac{1}{1426} = 0.0001957$ at the county level, $\rho = 0.00063$ at the state level, and $\rho = 0.000455$ at the national level.

We consider applying the Gaussian mechanism to satisfy differential privacy on children population estimate data along with a hierarchical constraint that ensures that the sum of the noisy population estimate of sub-hierarchies are consistent with their parent. For example, the sum of Title I-eligible children in the county-level subdivisions of Illinois should be consistent with the noisy estimate of Title I-eligible children at the state level. We experiment with three privacy budgets $\rho = 2.65$ and $\rho = 1.0$ and $\rho = 0.1$ to release privately the schools' population. These private counts are used to determine amount of money each school district should receive

using the allocation mechanisms described above for the basic, concentration, and target grants. The statistical bias in Definition 5 here represents the gain/loss in USD a school can have under the privacy preserving mechanism. 3 summarizes our findings of the bias as a function of district size for all 13,190 districts. We observe that misallocations are most pronounced at thresholds for each grant, and substantially impact smaller school districts more than larger ones.

It can be seen from Figure 3 that schools of low population receive far more money than they actually need as a consequence of the bias in the allocation mechanism, while larger districts may actually receive less funding. This figure made a clear evidence on the disparate impact of the privacy preserving mechanism in practice.

Analysis. In order to probe the effects of applying differential privacy to the Title I Allocation problem, we simulate applying DP and recomputing grants for each district based on the distribution formula described by (?). We compute 100K trials of samples to construct figures 3 and the rest of our analysis. We attribute the complex funding misallocation outcomes to the interaction of the several thresholds that constitute each grant.

An in-depth analysis of the allocation of Title I education funds in districts requires an examination of both total misallocations and misallocations per student. The consideration of these two perspectives is crucial for understanding the impact of differential privacy on the distribution of funds and its implications for educational equity.

Total misallocations offer a comprehensive overview by capturing the absolute magnitude of funding disparities across states. By assessing the overall amount of misallocated funds, we can identify states that are disproportionately affected. Such significant total misallocations can result in substantial disparities in educational resources, thereby hindering the ability of disadvantaged districts to provide quality education. The findings from 3, which displays the data on total misallocations for each type of grant (basic, concentration, and targeted), reaffirm this observation.

However, evaluating misallocations per student provides a more nuanced understanding of the fairness of funding distribution. This perspective allows us to account for the relative impact on individual students within each state . Misallocations per student highlight the distributional challenges faced by specific states and reveal the extent to which educational inequalities are exacerbated. The data presented in Figure 4, which depicts misallocations per student, demonstrate that smaller districts tend to gain a substantial amount of funds due to the application of differential privacy. In contrast, larger districts experience significant misallocations without necessarily facing the same level of negative impact on a per-student basis.

It is important to recognize the specific implications of misallocations in different contexts. When funds are misallocated in total, certain types of funding, such as specialized educational programs, additional resources, or support services for at-risk students, may be inadequately supported. This can perpetuate existing disparities and hinder efforts to address educational inequities, particularly in states with a higher population of disadvantaged students. On the other hand, misallocations per student shed light on the deprivation of financial support individual students require

to succeed academically. This impacts the availability of crucial resources, including instructional materials, qualified teachers, technology, and extracurricular activities, thereby impeding educational progress and perpetuating systemic inequalities.

The correlation between a state's student population and the alpha fairness, as shown in 6, further supports the observed disparities. The positive Pearson correlation coefficient of 0.76 indicates that states with larger populations of students tend to have higher levels of alpha fairness, indicating a greater degree of inequitable fund allocation.

Thus, the analysis of both total misallocations and misallocations per student provides a comprehensive understanding of the impact of differential privacy on the allocation of Title I education funds. This dual perspective enables policymakers to identify states disproportionately affected in terms of both the absolute amount of misallocated funds and the relative impact on individual students. By gaining insights into the specific challenges faced by different states, policymakers can devise targeted strategies to ensure a fair and equitable distribution of education funds, taking into account the disparities arising from both large total misallocations in large districts and smaller misallocations per student in small districts (Figure 5).

Section 203 of the Voting Rights Act

Section 203 of the Voting Rights Act requires certain jurisdictions to provide bilingual election materials and assistance to voters who are not proficient in English. To determine which jurisdictions are covered by this provision, the Census Bureau collects data on the number and percentage of voting-age citizens who are members of a language minority group and have limited English proficiency.

The Census Bureau uses three measures to calculate these numbers: the Limited English Proficient Population Count (LEPPCT), the Illiteracy Rate (ILLRAT), and the Voting Age Citizen Language Minority Group Citizen Voting Age Population (VACLEP).

The LEPPCT is the percentage of people in a language minority group who have limited English proficiency. The ILLRAT is the rate of total LMG voting age citizens who are limited-English proficient and have less than a 5th grade education. The VACLEP is the number of voting-age citizens who are members of a language minority group.

By using these three measures, the Census Bureau determines which jurisdictions meet the criteria for coverage under Section 203. This information is then used by election officials to provide bilingual election materials and assistance to voters who need it, ensuring that everyone has an equal opportunity to participate in the electoral process.

Allocation Formula. Coverage from Section 203 is determined by a simple mechanism that ensures that a county satisfies the following conditions:

The ILLRAT is greater than the national illiteracy rate AND the LEPPCT is greater than 5% OR the VACLEP is greater than 10,000. There are further nuances in this computation for American Indian and Alaska Native areas which we will refer to (10).

Analysis. In figure 10, we observe the effects of applying differential privacy to the population counts in a similar way

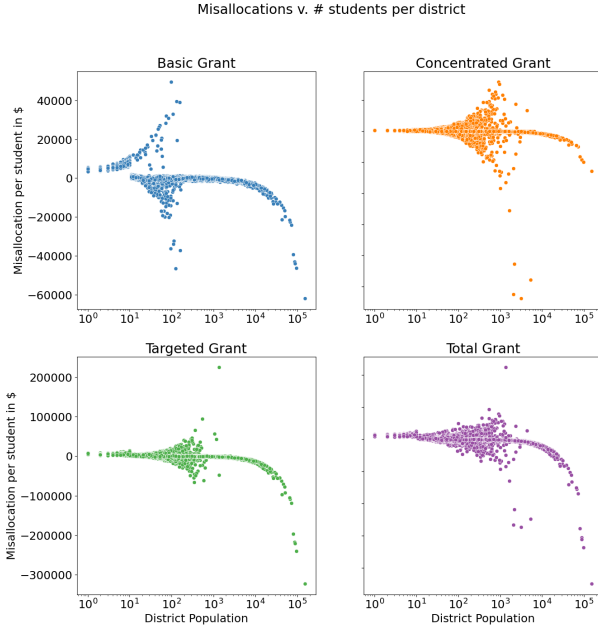


Fig. 3. Misallocations in USD compared to district size

Fig. 5. We apply differential privacy with a privacy budget of $\rho = 2.65$ to the population and eligible population counts of each district. We then plot the Title I misallocations for each district in the US for the basic, concentrated, and targeted grants, calculated using the allocation formulas outlined in (9). We observe that while large districts are more likely to receive larger misallocations, smaller districts are more likely to receive larger misallocations per student. We attribute these results to the complex interaction between differential privacy and the thresholds in the allocation formulas.

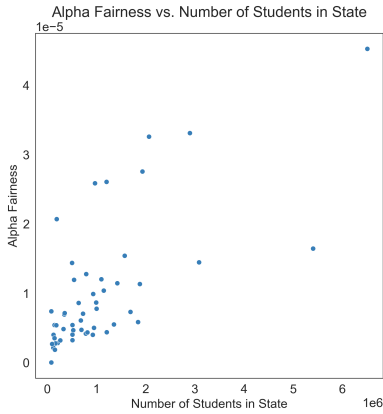


Fig. 6. Correlation between number of students per state and alpha fairness at $\rho = 2.65$

to our approach in the Title I misallocations. We observe that applying DP noise can lead to false negative outcomes, where communities that would originally receive coverage from Section 203 would be denied because their noisy LEPPCT value would not satisfy the threshold constraints

Differentially Private mechanisms

To release privately the outcomes P_i^F given a privacy constraint ϵ , there are several mechanisms. These mechanisms can roughly be divided into the two categories, strict and non-strict allocation mechanisms. A strict allocation mechanism requires that its outcome should always lie in the probability

Misallocations per student v. # students per district

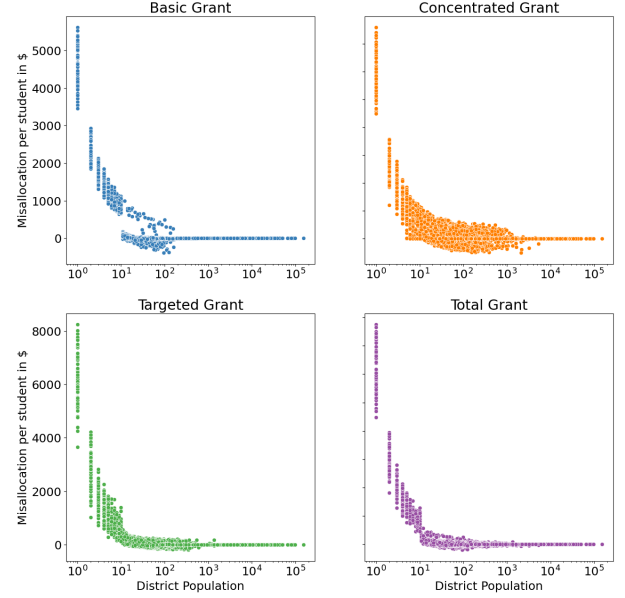


Fig. 4. Misallocations in USD per student compared to district size

simplex $\Delta_n = \{x \mid x \in \mathbb{R}^n, \mathbf{1}^T x = 1\}$ while a non-strict allocation mechanism only asks its output to be non-negative. The rest of this report aims to study the (approximate) optimal (strict allocation) mechanisms under different fairness metrics.

Strict Allocation Mechanism.

Definition 8 (Baseline Mechanism (BL)) The baseline mechanism outputs the allocation for each distribute $i \in [n]$ as follows.

$$\mathcal{M}_{\text{BL}}(\tilde{x})_i = \frac{a_i \cdot (\tilde{x}_i)_+}{\sum_{j=1}^n a_j \cdot (\tilde{x}_j)_+}.$$

Where \tilde{x}_i is the noisy private population count, while the superscript $x_+ = \max(x, 0)$ takes the non-negative part of the number x .

Definition 9 (Projection onto Simplex Mechanism (PoS))

The projection onto simplex mechanism outputs the allocation for each distribute $i \in [n]$ as follows.

$$\mathcal{M}_{\text{PoS}}(\tilde{x})_i = \arg \min_{v \in \mathbb{R}^n} \|v - P^F(\tilde{x})\|_2 \quad \text{s.t.} \quad \sum_{i=1}^n v_i = 1, v \geq 0$$

Non-strict Allocation Mechanism.

Definition 10 (Positive Allocation Mechanism (PA))

The positive allocation mechanism outputs the allocation for each distribute $i \in [n]$ as follows.

$$\mathcal{M}_{\text{PA}}(\tilde{x})_i = (P_i^F(\tilde{x}))_+ = \left(\frac{a_i \cdot \tilde{x}_i}{\sum_{j=1}^n a_j \cdot \tilde{x}_j} \right)_+.$$

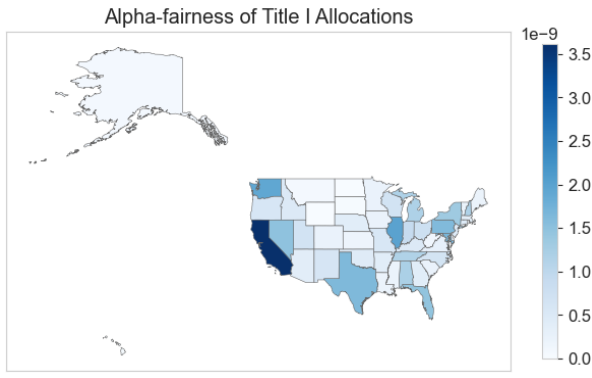


Fig. 7. Level of α fairness per state under Baseline mechanism (BL) at $\rho = 2.56$

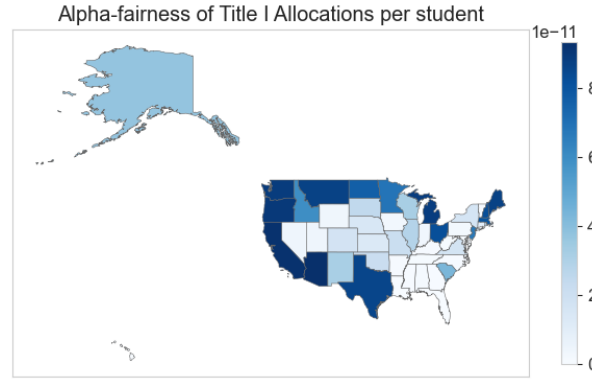


Fig. 8. Level of α fairness per student under Baseline mechanism (BL) at $\rho = 2.56$

Fig. 9. Fairness comparison

Definition 11 (Repair Mechanism (RP) (1)) The repair mechanism outputs the allocation for each distribute $i \in [n]$ as follows.

$$\mathcal{M}_{\text{RP}}(\tilde{x})_i = \frac{a_i \cdot (\tilde{x}_i)_+ + \Delta}{\sum_{j=1}^n a_j \cdot (\tilde{x}_j)_+ - \Delta'},$$

where

$$\Delta = \frac{\ln(2n/\delta)}{\epsilon}, \quad \Delta' = \frac{n \ln(2n^2/\delta)}{\epsilon}.$$

Proposition 1 (No-penalty allocation (1)) The following inequality holds with probability at least $1 - \delta$.

$$\mathcal{M}_{\text{RP}}(\tilde{x})_i \geq P_i^F(x), \quad \forall i \in [n].$$

Source of unfairness

We investigate the two main sources of unfairness highlighted in previous section: (1) shape of allocation function and (2) post-processing steps.

Shape of allocation function.

Theorem 1 Let P be an allotment problem which is at least twice differentiable. A data-release mechanism \mathcal{M} is α -fair w.r.t. P for some $\alpha < \infty$ if there exist some constant values $c_{j,l}^i$ ($i \in [n], j, l \in [k]$) such that, for all datasets $x \in \mathcal{X}$,

$$(HP_i)_{j,l}(x) = c_{j,l}^i \quad (i \in [n], j, l \in [k]).$$

Corollary 1 If P is a linear function, then \mathcal{M} is fair w.r.t. P .

Corollary 2 \mathcal{M} is fair w.r.t. P if there exists a constant c such that, for all dataset x ,

$$\text{Tr}(HP_i)(x) = c \quad (i \in [n]).$$

Corollary 3 Consider an allocation problem P . Mechanism \mathcal{M} is not fair w.r.t. P if there exist two entries $i, j \in [n]$ such that $\text{Tr}(HP_i)(x) \neq \text{Tr}(HP_j)(x)$ for some dataset x .

The above implies that fairness cannot be achieved if P is a non-convex function, as is the case for all the allocation problems considered in this paper. A fundamental consequence of this result is the recognition that adding Laplacian noise to the inputs of the motivating example will necessarily introduce fairness issues. For instance, consider P^F and notice that the trace of its Hessian

$$\text{Tr}(HP_i^F) = 2a_i \left[\frac{x_i \sum_{j \in [n]} a_j^2 - a_i \left(\sum_{j \in [n]} x_j a_j \right)}{\left(\sum_{j \in [n]} x_j a_j \right)^3} \right],$$

is not constant with respect to its inputs. Thus, any two entries i, j whose $x_i \neq x_j$ imply $\text{Tr}(HP_i^F) \neq \text{Tr}(HP_j^F)$. As illustrated in Figure ??, Problem P^F can introduce significant disparity errors. For $\epsilon = 0.001, 0.01$, and 0.1 the estimated fairness bounds are $0.003, 3 \times 10^{-5}$, and 1.2×10^{-6} respectively, which amount to an average misallocation of \$43,281, \$4,328, and \$865.6 respectively. The estimated fairness bounds were obtained by performing a linear search over all n school districts and selecting the maximal $\text{Tr}(HP_i^F)$.

Impact of post-processing. The post-processing steps can be applied at the inputs x , over the outcome $P_i^F(x)$ or at both. We investigate the impact of post-processing over input and outcome separately in this section.

Post-processing over the inputs. This step is performed to make sure the released private counts satisfies consistency constraints (11). For example, the released private counts should be non-negative integer numbers, or sum of counts at all cities' in a state should be equal to that state's count.

Non-negative truncation $\hat{x} = \max(0, \tilde{x})$

We have the following result which state that non-negative truncation introduces positive bias, and the closer to zero the true count is, the higher the bias.

Theorem 2 Let $\tilde{x} = x + \text{Lap}(\lambda)$, with scale $\lambda > 0$, and $\hat{x} = PP^{\geq \ell}(\tilde{x})$, with $\ell < x$, be its post-processed value. Then,

$$\mathbb{E}[\hat{x}] = x + \frac{\lambda}{2} \exp\left(\frac{\ell - x}{\lambda}\right).$$

Integral transform

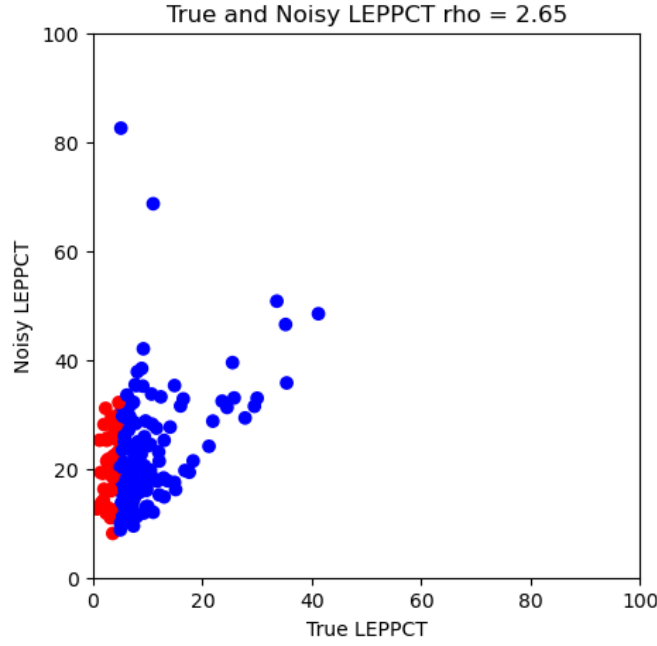


Fig. 10. LEPPCT vs. LEPPCT noisy values after DP mechanism at $\rho = 2.65$

The integral transform $PP^N(z)$ is used when the released data should be of integral quantities. To make sure that this processing step does not introduce additional bias, we can rely on the stochastic rounding technique:

$$PP^N(z) = \begin{cases} \lfloor z \rfloor & \text{w.p.: } 1 - (z - \lfloor z \rfloor) \\ \lfloor z \rfloor + 1 & \text{w.p.: } z - \lfloor z \rfloor \end{cases} \quad [5]$$

The stochastic rounding guarantees that $\mathbb{E}[PP^N(\tilde{x})] = \tilde{x}$ so no additional bias will introduce to $PP^N(\tilde{x})$

Post-processing over the outcomes.

Mitigating Solutions

Mechanisms. Different kind of post-processing mechanisms are considered in this section. These mechanisms require that their outcomes should always lie in the probability simplex Δ_n . The rest of this work aims to study the (approximate) optimal mechanisms under different fairness metrics.

Definition 12 (Baseline Mechanism (BL)) The baseline mechanism outputs the allocation for each entity $i \in [n]$ as follows.

$$\mathcal{M}_{BL}(\tilde{x})_i = \frac{a_i \cdot (\tilde{x}_i)_+}{\sum_{j=1}^n a_j \cdot (\tilde{x}_j)_+}.$$

Definition 13 (Projection onto Simplex Mechanism (PoS)) The projection onto simplex mechanism outputs the allocation for each entity $i \in [n]$ as follows.

$$\mathcal{M}_{PoS}(\tilde{x})_i = \arg \min_{v \in \mathbb{R}^n} \|v - P^F(\tilde{x})\|_2 \quad \text{s.t.} \quad \sum_{i=1}^n v_i = 1, v \geq 0$$

ACKNOWLEDGMENTS. Please include your acknowledgments here, set in a single paragraph. Please do not include any acknowledgments in the Supporting Information, or anywhere else in the manuscript.

1. S Kuppam, et al., Fair decision making using privacy-protected data (2020).
2. L Simunaci, Pro rata vaccine distribution is fair, equitable (t.ly/sDa9) (2021).

3. Title 13, Title 13, u.s. code (www.census.gov/history/www/reference/privacy_confidentiality/title_13_us_code.html) (2006) Accessed: 2021-01-15.
4. GDPR, What is gdpr, the eu's new data protection law? (<https://gdpr.eu/what-is-gdpr>) (2020) Accessed: 2021-01-15.
5. JM Abowd, The us census bureau adopts differential privacy in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 2867–2867 (2018).
6. C Dwork, F McSherry, K Nissim, A Smith, Calibrating noise to sensitivity in private data analysis in *Theory of cryptography conference*. (Springer), pp. 265–284 (2006).
7. C Dwork, A Roth, The algorithmic foundations of differential privacy. *Theor. Comput. Sci.* **9**, 211–407 (2013).
8. F Fioretto, P Van Hentenryck, Differential privacy of hierarchical census data: An optimization approach in *Proceedings of the International Conference on Principles and Practice of Constraint Programming (CP)*. pp. 639–655 (2019).
9. W Sonnenberg, Allocating grants for title i (2016).
10. RVRD Office, Section 203 language determinations (2022).
11. A Cohen, M Duchin, J Matthews, B Suwal, Census topdown: The impacts of differential privacy on redistricting in *2nd Symposium on Foundations of Responsible Computing*

869	(FORC 2021). (Schloss Dagstuhl-Leibniz-Zentrum für Informatik), (2021).	931
870		932
871		933
872		934
873		935
874		936
875		937
876		938
877		939
878		940
879		941
880		942
881		943
882		944
883		945
884		946
885		947
886		948
887		949
888		950
889		951
890		952
891		953
892		954
893		955
894		956
895		957
896		958
897		959
898		960
899		961
900		962
901		963
902		964
903		965
904		966
905		967
906		968
907		969
908		970
909		971
910		972
911		973
912		974
913		975
914		976
915		977
916		978
917		979
918		980
919		981
920		982
921		983
922		984
923		985
924		986
925		987
926		988
927		989
928		990
929		991
930		992