Population Research and Policy Review (2025) 44:8
https://doi.org/10.1007/s11113-024-09931-1

ORIGINAL RESEARCH

Protecting Vulnerable Respondents: A Critical Analysis of the Privacy-Preserving Methods of the 2010 and 2020 Decennial Census

Krishnamurty Muralidhar[1] · Josep Domingo-Ferrer[2] · David Sánchez[2] · Steven Ruggles[3]

Received: 20 November 2023 / Accepted: 16 December 2024

Abstract
This paper offers a comprehensive analysis of the statistical disclosure limitation (SDL) methodologies employed by the U.S. Census Bureau on the 2010 and 2020 Decennial Census releases under the perspective of the disclosure risk of the most vulnerable respondents. We first review the SDL methodology used up to the Decennial Census 2010, which was based on targeted swapping. Second, we examine recently reported reconstruction and reidentification results on the Decennial Census 2010 outputs, which form the foundation for the U.S. Census Bureau's decision to switch to a differentially private (DP) method for the 2020 release. Third, we examine the actual privacy and data accuracy achieved by the DP method and compare with the privacy and accuracy offered by the formerly employed swapping mechanism. We conclude that the DP method is not an adequate solution to protect the typically sparse tables present in the Decennial Censuses because it does not offer meaningful privacy guarantees in general, it poorly protects the privacy of the most vulnerable respondents in particular, and it significantly degrades the quality of the released data. We also argue that the claimed disclosure risks of previous Census releases were overstated because of a flawed reidentification procedure. Therefore, the U.S. Census Bureau's decision to change the SDL methodology to a DP-based one for the 2020 release was not only unwarranted, but it also reduced privacy and data quality compared to former releases.

Keywords Official statistics · U.S. census · Statistical disclosure limitation · Differential privacy · Data swapping

Extended author information available on the last page of the article

Published online: 09 January 2025

Introduction

In a recent article, Keller and Abowd (2023), respectively the current Chief Scientist at the U.S. Census Bureau and her immediate predecessor, state regarding the reidentification risk of the tabular data released from the 2010 Census: "It is only a matter of time before a malicious actor replicates the types of reconstruction attacks that Dick et al. (2023) and the U.S. Census Bureau have demonstrated." Starting around 2018, similar claims have been made in multiple presentations and articles (Abowd, 2018).

The 2010 Decennial Census data were protected with a statistical disclosure limitation (SDL) methodology based on targeted swapping. What was considered safe in 2010 has thereafter been put into question mainly by the U.S. Census Bureau itself. However, despite these predictions about using reconstruction attacks to reidentify individuals, to date there has not been a single confirmed case of reidentification of a respondent from the 2010 tabular data release. Nevertheless, these alleged reidentification risks form the foundation for the U.S. Census Bureau's decision to change the swapping-based methodology used in the 2010 Decennial Census to a differentially private method for the 2020 release. As a precursor to this, the U.S. Census Bureau has released many versions of the Disclosure Avoidance System where differential privacy was applied to the 2010 data. The release of the Disclosure Avoidance System has caused consternation among the users of Census due to the poor quality of the released data (Santos-Lazoda et al., 2020; Dove, 2021; Kenny et al., 2021; Hotz et al., 2022; Ruggles & Van Riper, 2021; Schneider, 2022; National Academies of Sciences, Engineering, and Medicine, 2023; Kenny et al., 2024), an issue that was not raised for the original 2010 release protected via targeted swapping. Given that the Census data are used not only by researchers but also by policy makers in important decisions, the significant data quality loss of the 2020 vs. 2010 releases is certainly worrying.

In this paper, we address two important questions: (1) *Was the U.S. Census Bureau justified in concluding that record reconstruction and subsequent reidentification results were enough to warrant a change from the swapping-based methodology used in the 2010 decennial Census?* (2) *Compared to the old method, does the new method offer the same level of protection to Census respondents who are considered to be especially vulnerable to disclosure?*

Background

In 2010, the collection of responses by household and then by individuals within household represented the raw Census data. These data were first edited to remove all errors and other issues. The identifying information was then removed from the file and replaced with a secret, unique Protected Identification Key (PIK). The resulting file is the Census Edited File (CEF), which is the source of all the information that is released.

Title 13 of the United States Code[1] prevents the U.S. Census Bureau from releasing data that could lead to identifying an individual from the released data. Since releasing data directly from CEF could potentially violate Title 13 requirements, the U.S. Census Bureau implemented statistical disclosure limitation (SDL) procedures on the CEF data. Zayatz et al. (2009) described the SDL procedures based on targeted data swapping adopted for tabular data release for the 2000 and 2010 Census as follows:

> "A small sample of households from the internal census data files was selected. Data from these households were swapped with data from other households that had identical characteristics on a certain set of variables but were from different geographic locations. Which households were swapped was not public information. The selection process was highly targeted to affect the records with the most disclosure risk. There was a threshold value for not swapping in blocks with a high imputation rate. Only records which were unique in their block based on a set of key demographic variables were swapped. The probability of being swapped had an inverse relationship with block size. In addition, records representing households containing members of a race category which appeared in no other household in that block had an additional p_1 probability of selection. All data products were created from the swapped file."

The result of applying this procedure to the CEF is called the Hundred Percent Detail file (HDF). *All released tabular data are based on HDF*.

According to the description above, the swapping methodology was targeted towards certain records that were defined at risk of disclosure (i.e., *vulnerable records*). Unfortunately, the U.S. Census Bureau does not release information on exactly how *vulnerable records* were determined. We can only assume that this was a *policy decision based on the interpretation of the requirements of Title 13*. Even if we do not know the exact details of which records were swapped and the reason (or criteria used) for swapping, there can be no question that this process required an in-depth analysis of the tabular data *prior to their release*.

To illustrate the U.S. Census Bureau's swapping approach, we reproduce the following anecdotal example reported in the New York Times (Hansen, 2018):

> "The bureau has long had procedures to protect respondents' confidentiality. For example, census data from 2010 showed that a single Asian couple — a 63-year-old man and a 58-year-old woman — lived on Liberty Island, at the base of the Statue of Liberty.
> That was news to David Luchsinger, who had taken the job as the superintendent for the national monument the year before. On Census Day in 2010, Mr. Luchsinger was 59, and his wife, Debra, was 49. In an interview, they said they had identified as white on the questionnaire, and they were the island's real occupants."

[1] https://www.govinfo.gov/content/pkg/USCODE-2007-title13/pdf/USCODE-2007-title13.pdf

In other words, the U.S. Census Bureau had identified the records of David and Debra Luchsinger as vulnerable to disclosure and swapped their records with that of an unknown couple changing their age and race. In SDL, being able to identify oneself in the data is referred to as *self-identification* and constitutes the strongest notion of reidentification risk. Therefore, reidentification is prevented if self-identification is prevented. In this case, Mr. and Mrs. Luchsinger can take comfort in the fact that *even if an adversary were to identify them using the Census tabular data based on their location, the adversary would not be able to get personal information about their age, race, and ethnicity*. We will return to this simple but important notion of self-identification later in this paper. But first, we discuss recent attempts to reconstruct individual-level microdata using tabular data.

Data and Methods

Reconstructing Individual-level Microdata using the 2010 Tabular Data Release

Starting around 2016 and motivated by the need to protect the forthcoming Decennial Census 2020, a team of U.S. Census Bureau researchers conducted experiments to evaluate the extent to which it was possible to reconstruct individual-level microdata from the tabular data released to the public from the 2010 Census. The details of these experiments can be found in Abowd (2021), Appendix B. This vulnerability was originally described by Dinur and Nissim (2003). In their study, they showed that, given a binary database protected by bounded additive noise of small magnitude relative to the size of the database, it may be possible for an adversary to accurately reconstruct the original data by issuing a random set of queries, creating a set of linear equations using the queries and responses, and solving these linear equations.

Whether the results obtained by Dinur and Nissim are directly applicable to the reconstruction of the Census tabular data is debatable. In the Dinur and Nissim (2003) study, the adversary is allowed to ask an unlimited number of queries where the *answer to the query is computed using the original data*, *random noise is generated and added to the true answer, and the response is computed as the true answer plus the random noise*. In traditional SDL, this is referred to as *output perturbation*. However, the tabular data release from the Census is very different. In this context, the original data are first modified via an SDL mechanism and all query responses are computed from the protected data. In addition, the adversary cannot issue queries to the database since the tabular data release corresponds to the responses to a pre-determined set of queries. In SDL, this is referred to as *input perturbation*. Hence, *the best an adversary can hope for is to reconstruct the protected data, since the original data are never accessed to respond to the queries*. Even Dinur and Nissim explicitly acknowledge this important difference. In their paper, they refer to a context like the one in the Census tabular data release as the "CD Model".

There have been three major studies that have addressed the issue of reconstructing individual-level microdata using the 2010 tabular data release, namely, the original reconstruction experiment by Abowd (2021), the subsequent modified

reconstruction by Hawes (2022), and another publication by U.S. Census Bureau researchers (Abowd et al., 2023). Additionally, Dick et al. (2023) also present a reconstruction effort that does not directly focus on reconstruction accuracy or reidentification (see Sánchez et al. (2023) and Sánchez et al. (2024)). Unfortunately, all these reconstructions attempts ignore the important difference that the tabular data released by the U.S. Census Bureau is based on the "CD Model" and is not based on queries. They follow the interactive query-based input perturbation approach of trying to set up and solve a series of linear equations, which results in dubious results and conclusions. More importantly, they raise claims on the ability of their methods to reconstruct *original* records, whereas what they can –probabilistically, rather than unequivocally– reconstruct are protected (masked) records.

In the most recent study related to reconstruction (Abowd et al., 2023), U.S. Census Bureau researchers make the claim that being able to reconstruct even the modified data constitutes disclosure. This seems to be an internal U.S. Census Bureau policy since there is no legal requirement to this effect. There is also no conceptual or theoretical reason as to why it constitutes disclosure since reconstruction only recreates the modified (protected) data and not the original (unprotected) data. In fact, according to Dinur & Nissim (2003), one of the greatest benefits of using the CD Model (where a modified, protected data is used to respond to queries) is that "users get a 'private' version of the database (written on a CD), which they may manipulate (say, without being restricted to statistical queries)." This statement makes it clear that reconstructing the modified (protected) database does not constitute a disclosure threat.

In addition, Muralidhar and Ruggles (2024) have also shown that an accurate reconstruction can be performed at the block level using fewer than a dozen tables released by U.S. Census Bureau. This reconstruction ability is *independent of the underlying protection mechanism* and is a function of the level of detail available in the tables. The solution to prevent the reconstruction of HDF is to change the level of detail in the tables either through suppression (for example, suppress P12A-P12I) or aggregation (for example, aggregating ages as under 18 and 18 and over, rather than the current 23 age groups). If the tables identified by Muralidhar and Ruggles (2024) are released with the same level of detail, reconstruction can be performed no matter whether the released data is based on swapping (as in the 2010 Census) or DP (as in the 2020 Census). Hence, the confidentiality failure of the 2010 Census described in Abowd et al. (2023) would also occur in the 2020 Census if the latter gave the same level of detail as the former.

Thus, in and of itself, reconstruction poses no threat if there is no corresponding reidentification. The critical aspect of the U.S. Census Bureau claims is that an accurate reconstruction leads to a high probability of reidentification.

Reidentifying Reconstructed Data

Reidentification was performed by the U.S. Census Bureau (Abowd, 2021) at the block level, initially, with individual year of age being reconstructed. Muralidhar (2022) has already addressed reidentification for the case of the individual year of

age and showed that it is practically random. Subsequently, the U.S. Census Bureau resorted to reidentification using "binned age" (age groups) rather than individual year of age (Hawes, 2022) and claimed that reconstructing by age group results in higher level of reconstruction and reidentification accuracy. Hence, in this section, we address reidentification by age group. Note that this choice should favor the reidentification claims made by the U.S. Census Bureau since using the age group (instead of the individual year of age) eliminates the uncertainty (and inaccuracy) resulting from dis-aggregating the age group to individual year of age during reconstruction.

To illustrate the error in the reidentification procedure described in Abowd (2021), in Table 1 we present data from three hypothetical blocks consisting of 10 male respondents in the 40 – 44 Age Group.

We assume that this reconstruction is accurate. We also assume that the external data source (consisting of Sex, Age, and Identification variables) is also accurate. Although it is highly unlikely, we will assume that the data for these three blocks have been released unmodified. We apply the procedure in Abowd (2021) with the exception that, as described in Hawes (2022), matching is performed based on the age group rather than the individual year of age. The results are as follows:

- In Block X, a match would be found for every individual and will result in the *confirmed reidentification of all the individuals in Block X*. The respondents in Block X are indistinguishable from one another and hence the initial assignment of identity (from the external source data) to the respondent is completely random. But whatever identity is assigned to whichever respondent in Block X, their identity is confirmed by the procedure in Abowd (2021). This contradicts the basic notion of hiding in a crowd (Chawla et al., 2005; Gehrke et al., 2012; Samarati, 2001).
- In Block Y, with probability 9/10 the Black respondent will be assigned a White identity, which means that a White respondent will be assigned a

Table 1 Three hypothetical blocks consisting of 10 male respondents of age 40–44

			Block X		Block Y		Block Z	
Respondent	Sex	Age group	Race	Ethnicity	Race	Ethnicity	Race	Ethnicity
1	M	(40–44)	White	Not Hispanic	Black	Not Hispanic	Black	Not Hispanic
2	M	(40–44)	White	Not Hispanic	White	Not Hispanic	Black	Hispanic
3	M	(40–44)	White	Not Hispanic	White	Not Hispanic	Asian	Not Hispanic
4	M	(40–44)	White	Not Hispanic	White	Not Hispanic	Asian	Hispanic
5	M	(40–44)	White	Not Hispanic	White	Not Hispanic	AIAN	Not Hispanic
6	M	(40–44)	White	Not Hispanic	White	Not Hispanic	AIAN	Hispanic
7	M	(40–44)	White	Not Hispanic	White	Not Hispanic	Other	Not Hispanic
8	M	(40–44)	White	Not Hispanic	White	Not Hispanic	Other	Hispanic
9	M	(40–44)	White	Not Hispanic	White	Not Hispanic	White	Not Hispanic
10	M	(40–44)	White	Not Hispanic	White	Not Hispanic	White	Hispanic

Black identity. In this case the identity of only 8 out of 10 individuals will be confirmed. Hence, the expected number of identity confirmations is 8*(9/10) + 10*(1/10) = 8.2 when the procedures in Abowd (2021) and Hawes (2022) are used on Block Y. Interestingly, the identity of the unique Black individual in this block is confirmed with much lower probability (10%) than the expected identity confirmation probability in the block (82%).

- Regarding Block Z, even though every individual is unique, with the procedure in Abowd (2021), *reidentification occurs purely by chance* (the probability of success is 1/10).

Thus, the procedure by Abowd (2021) yields the paradoxical result that unique individuals are less easily reidentified than homogenous individuals. U.S. Census Bureau's own Research and Methodology Directorate has suggested a procedure to overcome the problem of random identity assignment that explains the paradox. This procedure is described in McKenna (2019) as follows (underlining added):

> "For microdata, such reidentification studies are performed by looking for unique combinations of variables in the microdata that are thought to be identifying, looking for externally available data sets that contain the same variables, and then linking data records in the two data sets using the linkage variables. Finally, it is necessary to verify the proposed matches by comparing the suppressed identities in the microdata with the identities in the external data set to see if the matches are true matches or false matches. This last comparison step is vital, because often survey records are unique within the sample but not in the population (Ramachandran et al., 2012)."

If this procedure is implemented, since matching is performed only based on Sex and Age, and since no unique matches can be found on the matching variables, the probability of confirmed reidentification in Block X would be 10%, exactly what would occur by chance alone.

Whereas homogeneity makes *correct* reidentification more difficult, it facilitates reconstruction. Block X is homogenous, since every respondent has the same Sex, Age, Race, and Ethnicity, and the matching based on Sex and Age results in 100% reconstruction accuracy. Ruggles & Van Riper (2021) also reached the same conclusion. They also argue that, when the blocks are mostly homogenous, reconstruction accuracy is always very high. We can easily derive a lower bound for reconstruction accuracy in mostly homogenous blocks. Consider a block where p represents the proportion of majority individuals with the same Race and Ethnicity, q represents the proportion of all others, and $p > q$. In the worst case for reconstruction, assume that every minority individual is incorrectly labeled as a majority individual (which also means that a proportion q of the majority individuals will be identified as

minority). The minimum reconstruction accuracy in this case is $1 - 2q$. As the block becomes more homogenous (q gets smaller), this minimum accuracy increases.[2]

The notion of self-identification is quite useful to assess reidentification. With self-identification, there is no need for an external source of data. We assume that the respondents from the block attempt to identify themselves in the reconstructed data. And since such reidentification can be performed by any respondent in the database, we can use this approach to evaluate the entire database. In addition, with self-identification, we must assume that the respondent *knows the true values of all the attributes*, and the identity of a respondent is assumed to be confirmed when the values of *all the attributes of the individual match one and only one respondent* in the data. Hence, *every self-identification is a confirmed reidentification.* Thus, self-identification represents the strongest measure of reidentification risk.

Assessing self-identification in the hypothetical blocks produces the following results:

(1) Since every respondent in Block X has the same Age, Sex, Race, and Ethnicity, reidentification occurs only by chance,
(2) Since there is a single black respondent in Block Y, this respondent is reidentified; for the other respondents, reidentification occurs only by chance.
(3) Since every respondent in Block Z is unique, all the respondents are reidentified.

It is important to note that these results are *the opposite of what was observed using the procedure described in* Abowd (2021). That unique individuals in a data set can be reidentified if their data are released unmodified is a cornerstone of all statistical disclosure limitation procedures (including differential privacy). The self-identification procedure confirms this, but the procedure in Abowd (2021) contradicts it.

Implementing this procedure for the 2010 tabular data release is easy. The self-identification procedure is performed for every record in CEF, that is, the untouched original data. The number of records correctly reidentified represents the true risk of reidentification. Note that the CEF represents the best-quality external source data that is available to the adversary (Abowd, 2021). Hence, this assessment represents a strong upper bound on the confirmed reidentification risk.

In summary, our analysis above shows that the reidentification claims made by Abowd (2021) are overstated because of the incorrect procedure adopted to assess confirmed reidentification. Unfortunately, over the last few years, this claim has been repeated in many presentations, court documents, and the popular media. As a result, the incorrect claim "a majority of the respondents to the 2010 Census can be reidentified" is now taken as fact and it was used to justify the decision to move from the swapping-based disclosure limitation technique employed in 2010 to the

[2] This is true even in the case of the Dinur and Nissim study. Consider the scenario where the count of zeroes (n_0) is much greater than the count of ones (n_1), that is, the data are homogenous. In this case, the response to a single query "Count of ones" can be used to accurately reconstruct the data.

new procedure based on differential privacy in 2020. Given the above, and regarding the first question set out in the introduction, *this decision was unwarranted*.

The Differentially Private Method

Differential privacy (DP) is a privacy model originally designed to anonymize responses to interactive queries on an unreleased database. DP achieves privacy by making the presence or absence of any single record in the database not noticeable in the query responses, up to a factor exponential in ε (Dwork, 2006). If each record corresponds to a different individual, this means that the individual's information stays confidential. Formally, the *response* of a query satisfies ε-DP if, for all databases D and D' differing in one record, it holds that

$$\frac{P(Response|D)}{P(Response|D\prime)} \leq e^{\varepsilon}.$$

The main benefit of using DP is that it offers this "mathematical privacy guarantee". However, the guarantee implies real protection only for small values of ε (i.e., $\varepsilon < 1$ (Dwork, 2011)); for larger ε, an individual's data may become noticeable in the query response with exponentially high probability.

A detailed description of the DP-based method employed by the U.S. Census Bureau is provided in Abowd and Hawes (2023). The most recent value of ε chosen by the U.S. Census Bureau for the 2020 tabular data release is an inappropriately large 52.83 (U.S. Census Bureau, 2023). By interpreting the formal notion of DP as an odds ratio, this is the equivalent of 87,857,224,421,453,458,734,648 to 1. Suffice it to say that this mathematical privacy guarantee does not protect anyone's privacy. When Apple employed DP to protect user data collection with $\varepsilon = 14$ (Tang et al., 2017), one of the original developers of DP, Frank McSherry, said the following: "Using an epsilon value of 14 per day strikes me as relatively pointless" (Greenberg, 2017). The privacy guarantee offered by the U.S. Census Bureau (with ε=52.83) is *more than 73,000 trillion times worse*. As a result, the practical re-identifiability of the DP-protected data is likely to be worse than with the swapping method (according to Tables 9 and 10 vs. 15–17 in Abowd et al., 2023).

And yet, the 2010 tables released under the DP method have been criticized for lack of accuracy, an issue that did not arise for the original 2010 release protected via swapping. The U.S. Census Bureau has gone through multiple iterations of generating these DP-protected data. But even in the most recent iteration, users have pointed to multiple logical inconsistencies resulting from the used DP-based method (Wines 2022; Cornell Program on Applied Demographics, 2023; National Academies of Sciences, Engineering, and Medicine, 2023; Vink, 2023), such as:

(1) Blocks with population that is entirely under water,
(2) Blocks with zero household and non-zero population,
(3) Blocks with more households than population,
(4) Blocks with only children,
(5) Unusable age distributions.

These logical inconsistencies can be attributed primarily to two factors:

(1) With the DP-based method, noise is being added to a very sparse table. According to Jarmin et al. (2023), there are over 161 billion cells resulting from the combination (Block × Sex × Age × Race × Ethnicity) of which, at most, only about 330 million (total population in the 2020 census) cells (about 0.2%) have non-zero values. Hence, much of the noise is added to cells that need no protection.
(2) With the DP-based method, it is also necessary that noise be added *independently* to individual and housing demographic tables. As a result, even simple summary statistics, like the total number of persons in a block, are different between individual and housing tables. Most of the inconsistencies listed above can be attributed to this.

The overall impact of the implementation of the DP-based method is that the quality of block-level data is so poor that the U.S. Census Bureau (2021) has acknowledged that 2020 block-level tabular release should not be used for any meaningful analysis.

Discussion

Evaluating Differential Privacy and Data Swapping Under the Same Standards

To understand the poor data accuracy and the lack of privacy guarantees discussed above, in the following we compare the DP-based method with the formerly employed data swapping mechanism. There are few studies comparing both approaches. One such analysis performed by the U.S. Census Bureau is described in Hawes and Rodríguez (2021). The second study is by Christ et al. (2022), where swapping was implemented either by selecting the rows to be swapped randomly or selecting the rows based on their similarity.

One key difference between swapping as implemented for the 2010 Census and the two studies above is that the 2010 Census swapping was *targeted*. This means swapping focused primarily on vulnerable records, especially households on small blocks or with a unique set of characteristics (McKenna, 2018). Swapping a "safe" record serves little purpose in terms of disclosure control. Hawes & Rodriguez (2021) indicate that, in their experiments, they used swap rates, "up to 100% if necessary". But in most cases, swapping beyond a certain proportion serves no purpose at all in preventing disclosure (think of a block where one household is swapped out only to be replaced by another *identical* household from a different block). Thus, compared to targeted swapping, a random swap offers little marginal benefit but results in a significant loss of accuracy.

Another interesting feature of the implementation in the 2010 Census is that entire households are swapped (Zayatz et al., 2009). U.S. Census Bureau chooses to swap entire households specifically to enhance utility. There is no reason that

swapping must be performed in this manner. In their seminal article, Dalenius and Reiss (1982) describe swapping as being performed on an attribute-by-attribute basis. Keller and Abowd (2023) indicate that over 44% of the U.S. population have a unique combination of sex and age at the block level. If necessary, these records could be easily protected using univariate swapping. Only the U.S. Census Bureau can evaluate whether this is necessary.

With the DP-based method, we assume the same state-level invariants with the additional common-sense requirements that all cell counts must be integer and non-negative. To directly compare data swapping and the DP-based method *using comparable standards*, it is necessary to make the following changes to the 2010 swapping procedure:

(1) Block-level invariants must be removed and replaced by state-level invariants, and
(2) If necessary, swapping must be allowed to be implemented on an attribute-by-attribute basis.

The most important aspect of these changes is that, once implemented, they allow us to directly compare swapping and the DP-based method at the block level. For simplicity, consider the scenario where the only invariant is to preserve the population total at the state level. With the DP-based method, this amounts to adding noise to the cells of the comprehensive frequency table containing one cell for each possible combination of attribute values. In this case, noise addition results in an increment in a cell count (the one corresponding to the new combination resulting from noise addition) that is balanced by a decrement in another cell count (the one corresponding to the original combination of values). *This is precisely what swapping does as well*.

Alternatively, the comparison can also be performed where the DP-based method is implemented to preserve the same block-level invariants as the released 2010 Census data. According to Dajani et al. (2017), the U.S. Census Bureau reached an agreement with the US Justice Department to preserve the total, voting age, and under 18 population counts invariant at the block-level. This was later changed to preserve *only the total population count* invariant only at the state-level (Abowd et al., 2020). The U.S. Census Bureau has *never* released demonstration data with block-level invariants. It is unclear whether preserving population counts invariant at the block-level is even possible with a DP-based method.

Protecting Vulnerable Records: Differential Privacy vs. Swapping

The difference between swapping and the DP-based method on the 2010 Census can be attributed exclusively to two choices: how the cells/records to be modified are selected and the probability that a particular cell/record will be modified. In terms of selection of records to be swapped, *the U.S. Census Bureau swapping approach is targeted while the DP method is random*. The description by McKenna (2018) indicates that the records to be swapped were identified by analyzing household

and individual-level data. Once identified, these records were swapped with another record.

We return to the earlier example from Liberty Island to illustrate this approach. This block consisted of only two individuals who were both unique. According to Title 13, the U.S. Census Bureau is prevented from releasing data that would enable the identification of an individual. Without disclosure limitation, the two individuals in the example would be readily (self) identified. Thus, to satisfy Title 13, these two vulnerable records *must* be modified. Hence, these two records were swapped with two other records prior to the tabular data release. Even if "there are *millions* of such records" (Groshen and Goroff, (2022), their emphasis), they can be protected in a similar manner. Given the total population of the 2010 U.S. population was over 307 million, *millions* could still be a small percentage of the population. Even if this percentage was larger, it can be argued that this would be the cost of disclosure prevention.

With the DP-based method, random (positive or negative) noise is added to all the cells in this block. According to Abowd (2021), within a given block there are 28,980 cells in the comprehensive table at the block level. In the case of Liberty Island, of all these cells, only 2 cells (corresponding to the two vulnerable respondents) are non-zero while the remaining are zero cells. *To satisfy Title 13 requirement (prevent reidentification), it is necessary for the cell counts corresponding to these two vulnerable respondents to be modified*. Unfortunately, that the noise added *must not be based on the original value and must be random* is a prime directive in DP (Dwork, 2006).[3] Hotz et al. (2022) and Hotz and Salvo (2022) have also highlighted the fact that DP based methods do not consider the prior probability of disclosure, which is a necessary component of preventing disclosure as defined by Title 13. As a result, randomly adding noise to cells (with the restriction of integer noise and non-negative cell counts) implies that there is a nonzero probability that these a priori at-risk two cells are not modified, thereby violating Title 13 requirements. And this probability increases exponentially if a large ϵ is chosen. The value of ϵ (52.83) is so large that the DP ratio is no longer meaningful. Even the founders of DP have called against the use of large ϵ values, going as far as to say that such implementations offer privacy in name only (Dwork et al., 2019).

In general, to protect vulnerable records, it is necessary that block cell counts that contain the vulnerable records be modified. DP precludes adding noise to cells which have a high a priori risk of disclosure, but this is precisely what is required to satisfy the requirements of Title 13 (Hotz & Salvo, 2022; Hotz et al., 2022). We can now answer the second question stated in the introduction: *a record identified by the U.S. Census Bureau as vulnerable was offered protection in the 2010 Census, but not in the 2020 Census*.

Further, since data swapping identifies and protects vulnerable records while maintaining non-vulnerable ones untouched, the loss of accuracy resulting from this

[3] Note that this requirement applies to post-processing as well. Hence, it would be violated by any attempt to force modification of specific cells during the application of the Top-Down algorithm used by the USCB for the 2020 tabular data release.

data protection mechanism is kept at a minimum. DP noise, on the other hand, is blindly applied across all cells, which results in many records being unnecessarily modified even for large values of ε; that is, much of the protection budget is wasted on cells/records that need no protection. This results in poor data quality, which is manifested in the significant inaccuracies discussed above.

Poor data quality is a common issue when employing DP in scenarios different from the one DP was meant for, as it is the case for the non-interactive Census data releases (Blanco-Justicia et al., 2022; Domingo-Ferrer et al., 2021). Since enforcing DP requires adding noise proportional to the sensitivity of the query, using DP for individual data releases requires considering the most sensitive query, i.e., the identity query. Therefore, a large ε value is required to avoid adding too much noise that would result in nearly random data which unfortunately dilutes the privacy guarantee. The poor performance of DP in Census-like data releases is unsurprising because DP was never designed to be applied outside the (constrained) interactive scenario. Even though many researchers and practitioners insist on applying DP for all scenarios, conciliating privacy and data accuracy with DP outside the interactive setting is, by the very definition of DP, impractical (Domingo-Ferrer et al., 2021).

Finally, we wish to remark that no information is publicly available regarding the criteria used by the U.S. Census Bureau to identify vulnerable records. We are open to the idea that a higher level of protection would have required stricter criteria, a larger number of vulnerable records and, consequently, more swapping. Only U.S. Census Bureau has access to this information. It would have been a simple matter for the U.S. Census Bureau to produce, as a part of their Disclosure Avoidance System, *comparable output using the same standards for both targeted swapping and the DP-based method.*

Conclusions

We have analyzed and compared the SDL methodologies employed by the U.S. Census Bureau in the 2010 and 2020 Decennial Census releases. We have shown that the traditional targeted swapping is an effective method to reconcile the privacy of vulnerable records with data accuracy, and that the risk of reidentification alleged in recently proposed reconstruction attacks has been overstated due to an incorrect reidentification methodology. As a matter of fact, despite recent efforts, no single confirmed case of reidentification from former Census releases is known. Therefore, the urgent need to move to a new SDL method is hard to justify.

We have also shown that the new SDL method employed by the U.S. Census Bureau applies DP so loosely that no meaningful privacy guarantee is achieved. Despite using an unreasonably large ε (in an effort to limit the amount of noise added to the data), data accuracy is significantly degraded. DP adds noise to all cells in very large sparse tables, rather than just modifying vulnerable records, as the swapping method does. Such an untargeted modification produces unnecessary distortion. Perhaps more importantly, whereas swapping ensures that vulnerable records are modified, under DP with a very large ε there is a high probability that

those records will not be modified, thereby failing to achieve the deterministic privacy protection required by Title 13 for such records.

In 2010, Professor Stephen Fienberg, a well-known statistician, and a key contributor to the SDL literature, wrote (Fienberg et al., 2010):

> "Because differential privacy provides guarantees for the method and not for the specific data at hand, we do not believe the methodology is suitable for the type of large sparse tables often produced by statistics agencies and sampling organizations."

The U.S. Census Bureau implementation of a DP-based procedure for the 2020 decennial census confirm this observation. After multiple attempts, we now have a situation in which the resulting output does not satisfy the data accuracy requirement and may not even protect vulnerable records.

Acknowledgements Partial support to this work has been received from the European Commission (project H2020-871042 "SoBigData++"), the Government of Catalonia (ICREA Acadèmia Prizes to J. Domingo-Ferrer and to D. Sánchez, and grant 2021SGR-00115), MCIN/AEI/ https://doi.org/10.13039/501100011033 and "ERDF A way of making Europe" under grant PID2021-123637NB-I00 "CURLING", and the EU's NextGenerationEU/PRTR via INCIBE (project "HERMES" and INCIBE-URV cybersecurity chair).

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. European Commission,871042,Josep Domingo-Ferrer,Ministerio de Ciencia,Innovación y Universidades,PID2021-123637NB-I00,Josep Domingo-Ferrer,Instituto Nacional de Ciberseguridad,HERMES,Josep Domingo-Ferrer,INCIBE-URV cybersecurity chair,David Sanchez,ICREA,ICREA Academia,Josep Domingo-Ferrer,ICREA Academia,David Sanchez,Departament d'Innovació,Universitats i Empresa,Generalitat de Catalunya,2021SGR-00115, Josep Domingo-Ferrer

Data availability No data applies to this paper.

Declarations

Competing Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

Abowd, J.M. (2018). Staring down the database reconstruction theorem. *Joint Statistical Meeting*, Vancouver, BC, Canada. Retrieved July 18, 2023, from https://www.census.gov/content/dam/Census/newsroom/press-kits/2018/jsm/jsm-presentation-database-reconstruction.pdf

Abowd, J. M., Benedetto, G. L., Garfinkel, S. L., Dahl, S. A., Dajani, A. N., Graham, M., & Hawes, M. B. (2020, August). *The modernization of statistical disclosure limitation at the U.S. Census Bureau*. U.S. Census Bureau. Retrieved July 18, from 2023 https://www.census.gov/library/working-papers/2020/adrm/CED-WP-2020-009.html

Abowd, J. M. (2021, April 13). Declaration of John M. Abowd. *Case no. 3:21-CV-211-RAH-ECM-KCN*. U. S. District Court for the Middle District of Alabama (13 April 2021).

Abowd, J. M., Adams, T., Ashmead, R., Darais, D., Dey, S., Garfinkel, S. L., Goldschlag, N., Kifer, D., Leclerc, P., Lew, E., Moore, S., Rodríguez, R. A., Tadros, R. N., & Vilhuber, L. (2023) The 2010 Census confidentiality protections failed, here's how and why. *NBER Working Paper Series, working paper 31995*. Retrieved August 14, 2024, from https://www.nber.org/system/files/working_papers/w31995/w31995.pdf

Abowd, J. M., & Hawes, M. B. (2023). Confidentiality protection in the 2020 U.S. Census of population and housing. *Annual Review of Statistics and Its Application, 10*, 119–144.

National Academies of Sciences, Engineering, and Medicine. (2023) *Assessing the 2020 Census: Final report*. Washington DC: The National Academies Press.

Blanco-Justicia, A., Sánchez, D., Domingo-Ferrer, J., & Muralidhar, K. (2022). A critical review on the use (and misuse) of differential privacy in machine learning. *ACM Computing Surveys, 55*(8), 1–16.

U.S. Census Bureau. (2021). *Disclosure avoidance for the 2020 Census: An introduction*. U.S. Government Publishing Office, Washington DC, USA. Retrieved July 18, 2023, from https://www2.census.gov/library/publications/decennial/2020/2020-census-disclosure-avoidance-handbook.pdf

U.S. Census Bureau. (2023, April 3). *Privacy loss budget allocation*. Retrieved November 11, 2024, from https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/04-Demonstration_Data_Products_Suite/2023-04-03/2023-04-03_Privacy-Loss_Budget_Allocations.pdf

Chawla, S., Dwork, C., McSherry, F., Smith, A., & Wee, H. (2005). Towards privacy in public databases. In *Proceedings of Theory of Cryptography Conference – TCC 2005* (pp. 363–385).

Christ, M., Radway, S., & Bellovin, S. M. (2022). Differential privacy and swapping: Examining de-identification's impact on minority representation and privacy preservation in the U.S. Census. In *Proceedings of the 2022 IEEE Symposium on Security and Privacy* (pp. 457–472).

Cornell Program on Applied Demographics. (2023, March 23). *Census 2020 results. Data and analyses for New York from the data products as they are released over time by the U.S. Census Bureau*. Retrieved July 18, 2023, from https://pad.human.cornell.edu/census2020/index.cfm#das

Dajani, A. N., Lauger, A. D., Singer, P. E., Kifer, D., Reiter, J. P., Machanavajjhala, A., Garfinkel, S. L., Dahl, S. A., Graham, M., Karwa, V., Kim, H., Leclerc, P., Schmutte, I. M., Sexton, W. M., Vilhuber, L., & Abowd, J. M. (2017). *The modernization of statistical disclosure limitation at the U. S. Census Bureau*. Census Scientific Advisory Committee Meeting, Suitland MD, U.S. Retrieved November 10, 2024, from https://www2.census.gov/cac/sac/meetings/2017-09/statistical-disclosure-limitation.pdf

Dalenius, T., & Reiss, S. P. (1982). Data-swapping: A technique for disclosure control. *Journal of Statistical Planning and Inference, 6*(1), 73–85.

Dick, T., Dwork, C., Kearns, M., Liu, T., Roth, A., Vietri, G., & Wu, Z. S. (2023). Confidence-ranked reconstruction of census microdata from published statistics. *Proceedings of the National Academy of Sciences of the United States of America, 120*(8), e2218605120.

Dinur, I., & Nissim, J. (2003). Revealing information while preserving privacy. In *Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (pp. 202–210).

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

Dove, I. (2021). *Applying differential privacy protection to ONS mortality data. Pilot study*. Office for National Statistics U.K. Retrieved July 18, 2023, form https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/applyingdifferentialprivacyprotectiontoonsmortalitydatapilotstudy

Dwork, C. (2006). Differential privacy. In M. Bugliesi, B. Preneel, V. Sassone, I. Wegener (Eds.), *Proceedings of Automata, Languages and Programming – ICALP 2006* (pp. 1–12).

Dwork, C. (2011). A firm foundation for private data analysis. *Communications of the ACM, 54*(1), 86–95.

Dwork, C., Kohli, N., & Mulligan, D. (2019). Differential privacy in practice: Expose your epsilons! *Journal of Privacy and Confidentiality*. https://doi.org/10.29012/jpc.689

Fienberg, S. E., Rinaldo, A., & Yang, X. (2010). Differential privacy and the risk-utility tradeoff for multi-dimensional contingency tables. In J. Domingo-Ferrer, E. Magkos, (Eds.), *Proceedings of Privacy in Statistical Databases – PSD 2010* (pp. 187–199).

Gehrke, J., Hay, M., Lui, E., & Pass, R. (2012). Crowd-blending privacy. In *Proceedings of Advances in Cryptology - CRYPTO 2012* (pp. 479–496).

Greenberg, A. (2017, September 15). How one of Apple's key privacy safeguards falls short. *Wired*. Retrived July 18, 2023, from https://www.wired.com/story/apple-differential-privacy-shortcomings/

Groshen, E. L., & Goroff, D. (2022). Disclosure avoidance and the 2020 census: What do researchers need to know? *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.aed7f34f

Hansen, M. (2018, December 5). To reduce privacy risks, the Census plans to report less accurate data. *New York Times*. Retrieved July 18, 2023, from https://www.nytimes.com/2018/12/05/upshot/to-reduce-privacy-risks-the-census-plans-to-report-less-accurate-data.html?partner=slack&smid=sl-share

Hawes, M., & Rodríguez, R. A. (2021, May). *Determining the privacy-loss budget: Research into alternatives to differential privacy.* Census Scientific Advisory Committee. Retrieved July 18, 2023, from https://www2.census.gov/about/partners/cac/sac/meetings/2021-05/presentation-research-on-alternatives-to-differential-privacy.pdf

Hawes, M., (2022, September 29–30). *Reconstruction and reidentification of the demographic and housing characteristics file (DHC)*. Presentation to the Census Scientific Advisory Committee, Washington DC, USA. Retrieved July 18, 2023, form https://www2.census.gov/about/partners/cac/sac/meetings/2022-09/presentation-reconstruction-and-re-dentification-of-dhc-file.pdf

Hotz, V. J., Bollinger, C. R., Komarova, T. V., Manski, C. F., Moffitt, R. A., Nekipelov, D., Sojourner, A., & Spencer, B. D. (2022). Balancing data privacy and usability in the federal statistical system. *Proceedings of the National Academy of Sciences of the USA, 119*, e2104906119.

Hotz, V. J., & Salvo, J. (2022). A chronicle of the application of differential privacy to the 2020 Census. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.ff891fe5

Keller, S. A., & Abowd, J. M. (2023). Database reconstruction does compromise confidentiality. *Proceedings of the National Academy of Sciences USA, 120*(12), e2300976120.

Kenny, C., Kuriwaki, S., McCartan, C., Rosenman, E., Simko, T., & Imai, K. (2021). The use of differential privacy for census data and its impact on redistricting: The case of the 2020 US Census. *Science Advances, 7*, eabk3283.

Kenny, C., McCartan, C., Kuriwaki, S., Simko, T., & Imai, K. (2024). Evaluating bias and noise induced by the U.S. Census Bureau's privacy protection methods. *Science Advances, 10*(18), eadl2524.

McKenna, L. (2018). *Disclosure avoidance techniques used for the 1970 through 2010 Decennial Censuses of Population and Housing*. Technical Report. U.S. Census Bureau, Washington, DC, U.S. Retrieved July 18, 2023, from https://www2.census.gov/ces/wp/2018/CES-WP-18-47.pdf

McKenna, L. (2019). *U.S. Census Bureau reidentification techniques*. Research and Methodology Directorate, U.S. Census Bureau. Retrieved July 18, 2023, from https://www2.census.gov/adrm/CED/Papers/CY19/2019-04-Reidentification%20studies-20210331FinRed.pdf

Muralidhar, K. (2022). A re-examination of the Census Bureau reconstruction and reidentification attack. In J. Domingo-Ferrer, M. Laurent, (Eds.), *Proceedings of Privacy in Statistical Databases - PSD 2022* (pp. 312–323).

Muralidhar, K. & Ruggles, D. (2024). Escalation of commitment: A case study of the United States Census Bureau efforts to implement differential privacy for the 2020 Decennial Census. In J.Domingo-Ferrer, M. Önen (Eds.) *Proceedings of Privacy in Statistical Databases, PSD 2024* (pp. 393–402)– https://arxiv.org/abs/2407.15957

Ramachandran, A., Singh, L., Porter, E., & Nagle, F. (2012). Exploring re-identification risks in public domains. In *Proceedings of the 10th Annual International Conference on Privacy Security and Trust* (pp. 35–42).

Ruggles, S., & Van Riper, D. (2021). The role of chance in the Census Bureau database reconstruction experiment. *Population Research and Policy Review, 41*(3), 781–788.

Samarati, P. (2001). Protecting respondents identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering, 13*(6), 1010–1027.

Sánchez, D., Jebreel, N., Muralidhar, K., Domingo-Ferrer, J., & Blanco-Justicia, A. (2024). An examination of the alleged privacy threats of confidence-ranked reconstruction of Census microdata. In J.Domingo-Ferrer, M. Önen (Eds.). *Proceedings of Privacy in Statistical Databases – PSD 2024* (pp. 213–224).

Sánchez, D., Domingo-Ferrer, J., & Muralidhar, K. (2023). Confidence-ranked reconstruction of census records from aggregate statistics fails to capture privacy risks and re-identifiability. *Proceedings of the National Academy of Sciences U.S.A, 120*(8), e2303890120.

Santos-Lozada, A. R., Howard, J. T., & Verdery, A. M. (2020). How differential privacy will affect our understanding of health disparities in the United States. *Proceedings of the National Academy of Sciences of the U.S.A, 117*(24), 13405–13412.

Schneider, M. (2022, August 8). Researchers ask Census to stop controversial privacy method. *U.S. News*. Retrieved July 18, 2023, from https://www.usnews.com/news/business/articles/2022-08-08/researchers-ask-census-to-stop-controversial-privacy-method

Tang, J., Korolova, A., Bai, X., Wang, X., & Wang, X. (2017). *Privacy loss in Apple's implementation of differential privacy on MacOS 10.12*. arXiv Preprint. Retrieved July 18, 2023, from https://arxiv.org/abs/1709.02753

Vink, J. (2023). Usability of age distributions in the Disclosure Avoidance System demonstration data. In: *Applied Demography Conference*. Retrieved August 13, 2024, from https://pad.human.cornell.edu/presentations/slides/ADC2023JV/ADC2023JV.pdf

Wines, M. (2022, April 21). The 2020 Census suggests that people live underwater. There's a Reason. *The New York Times*. Retrieved July 18, 2023, from https://www.nytimes.com/2022/04/21/us/census-data-privacy-concerns.html

Zayatz, L., Lucero, J., Massell, P., & Ramanayake, A. (2009). Disclosure avoidance for Census 2010 and American Community Survey five-year tabular data products. *Census Bureau Research Report RRS2009–10* Retrieved July 18, 2023, from https://www.census.gov/content/dam/Census/library/working-papers/2009/adrm/rrs2009-10.pdf

Authors and Affiliations

Krishnamurty Muralidhar[1] · Josep Domingo-Ferrer[2] · David Sánchez[2] · Steven Ruggles[3]

✉ David Sánchez
david.sanchez@urv.cat

Krishnamurty Muralidhar
krishm@ou.edu

Josep Domingo-Ferrer
josep.domingo@urv.cat

Steven Ruggles
ruggles@umn.edu

[1] Department of Marketing and Supply Chain Management, Price College of Business, University of Oklahoma, Norman, OK 73019, USA

[2] Department of Computer Engineering and Mathematics, CYBERCAT-Center for Cybersecurity Research of Catalonia, Universitat Rovira I Virgili, Av. Països Catalans, 26, 43007 Tarragona, Catalonia, Spain

[3] Institute for Social Research and Data Innovation, University of Minnesota, Minneapolis, MN 55455, USA