

SafeTab-H: Disclosure Avoidance for the 2020 Census Detailed Demographic and Housing Characteristics File B (Detailed DHC-B)

William Sexton¹, Skye Berghel¹, Bayard Carlson¹, Sam Haney¹, Luke Hartman¹, Michael Hay¹, Ashwin Machanavajjhala¹, Gerome Miklau¹, Amritha Pai¹, Simran Rajpal¹, David Pujol¹, Ruchit Shrestha¹, and Daniel Simmons-Marengo¹

¹Tumult Labs

May 28, 2024

Abstract

This article describes SafeTab-H, a disclosure avoidance algorithm applied to the release of the U.S. Census Bureau’s Detailed Demographic and Housing Characteristics File B (Detailed DHC-B) as part of the 2020 Census. The tabulations contain household statistics about household type and tenure iterated by the householder’s detailed race, ethnicity, or American Indian and Alaska Native tribe and village at varying levels of geography. We describe the algorithmic strategy which is based on adding noise from a discrete Gaussian distribution and show that the algorithm satisfies a well-studied variant of differential privacy, called zero-concentrated differential privacy. We discuss how the implementation of the SafeTab-H codebase relies on the Tumult Analytics privacy library. We also describe the theoretical expected error properties of the algorithm and explore various aspects of its parameter tuning.

Contents

1	Introduction	3
2	Problem Setup	4
2.1	Households	4
2.2	Geography	4
2.3	Race and Ethnicity	5
2.4	Population Groups	6
2.5	Detailed Demographic and Housing Characteristics File B	7
2.6	Private Release Problem	10
3	SafeTab-H Algorithm	10
3.1	Input Data Description	11
3.1.1	Household Data	11
3.1.2	Total Population Counts	12
3.2	The Algorithm Description	12
4	Privacy Preliminaries	17
4.1	Privacy Definitions	17
4.2	Privacy Properties	17
4.2.1	Composition	17
4.2.2	Postprocessing	18
4.3	Base Mechanism	18
5	SafeTab-H Privacy Analysis	18
5.1	Converting to Bounded zCDP	19
6	Implementation of SafeTab-H	20
6.1	Input Validation	20
6.2	Tumult Analytics	20
6.3	Postprocessing for Addressing Demographic Reasonableness Concerns	21
6.3.1	Marginals	21
6.3.2	Suppression	22
6.3.3	Coterminous Geographies	22
6.3.4	Tabulation System Suppression	22
6.4	Input Sourcing	23
6.5	Other Implementation Details	23
6.5.1	Mapping Detailed Race and Ethnicity Codes to Characteristic Iterations	23
6.5.2	Puerto Rico	23
7	Parameters and Tuning	24
7.1	Error Bounds	24
7.2	Parameter Identification, Trade-offs, and Outcomes	25
7.3	Parameter Tuning Using the SafeTab-H Analysis Tool	25
7.3.1	Parameter Selection	25
8	Conclusion	26

1 Introduction

It is the responsibility of the U.S. Census Bureau (Census Bureau) to conduct a census of the U.S. population every 10 years. The 2020 Census is the latest of such efforts, which aims to enumerate every person living in the United States. As part of the 2020 Census undertaking, the Census Bureau manages 35 operations (e.g., Address Canvassing, Nonresponse Followup, and Redistricting Data Program) [1]. Each of these operations controls a number of systems that handle various aspects of the entire census endeavor, ranging from data collection and processing to the dissemination of data products to the U.S. people. Throughout these 2020 Census procedures, the Census Bureau strives to maintain the privacy and confidentiality of its respondents. The Disclosure Avoidance System (DAS) manages the confidentiality protection of statistical data releases from the 2020 Census. The DAS executes its duties after census responses have been collected and processed into a database known as the Census Edited File (CEF) but before data products are released for public consumption. For the 2020 Census, the DAS was redesigned to modernize its privacy protection mechanisms. The modernization effort provides individuals and households with state-of-the-art protection against the privacy threats associated with releasing census data to the public.

The Census Bureau releases several different data products for the 2020 Census, including the 2020 Census Redistricting Data (P.L. 94-171) Summary File, Demographic and Housing Characteristics File (DHC), Demographic Profile, Detailed Demographic and Housing Characteristics File A (Detailed DHC-A), Detailed Demographic and Housing Characteristics File B (Detailed DHC-B), and Supplemental Demographic and Housing Characteristics File (S-DHC). Each data product exhibits distinct challenges concerning confidentiality protection. Hence, the DAS deploys various algorithms to optimize protection and accuracy across each data product. We note that some data products share similar enough challenges to utilize the same algorithm. Despite algorithmic differences, all statistical disclosure limitation techniques fit into the same overarching privacy framework known as *differential privacy*. The differential privacy framework calls for the design of algorithms that satisfy mathematically provable guarantees regarding the data publication process. Section 4.1 gives further details. Although the differential privacy framework encompasses several different privacy definitions, in this paper, one should assume we use the term “differential privacy” to refer to zero-concentrated differential privacy unless otherwise specified.

SafeTab-P and SafeTab-H are two of the privacy algorithms deployed by the DAS. SafeTab-P provides privacy protection for the Detailed DHC-A whereas SafeTab-H was designed to provide differential privacy guarantees for the Detailed DHC-B. SafeTab-H is the primary topic of this paper, but we mention SafeTab-P for two reasons: (1) both algorithms share many similarities because the Detailed DHC-A and Detailed DHC-B are closely related data products and (2) some outputs of the SafeTab-P algorithm appear as inputs to the SafeTab-H algorithm. We refer the reader to Section 2 for an in-depth description of the data product and privacy release problem.

The main goals of this article are threefold:

1. Describe the SafeTab-H algorithm and how it meets the requirements of the Detailed DHC-B.
2. Prove the privacy properties of the SafeTab-H algorithm.
3. Describe the parameters in the SafeTab-H algorithm and how they impact privacy-accuracy trade-offs.

For the first point, we provide a technical pseudocode description of SafeTab-H that details how privacy protection is applied to create the Detailed DHC-B tabular summaries (Section 3).

We also highlight salient differences between our pseudocode abstraction and the programmed codebase of the algorithm (Section 6). While this article provides meaningful context regarding the implemented algorithm, it does not provide a detailed overview of the code architecture (e.g., module interactions and class descriptions). The SafeTab-H codebase is open source for interested readers.

For the second point, we provide relevant background material on the differential privacy framework and explain why SafeTab-H adheres to the framework (Section 5).

For the third point, we discuss the parameters in SafeTab-H that impact the privacy-accuracy trade-offs of the algorithm. We cover parameter tuning, including a brief look at related data accuracy considerations (Section 7).

2 Problem Setup

The Detailed DHC-B contains statistics (counts) of household type and tenure, including total household count, for households in the United States and Puerto Rico, crossed with detailed races and ethnicities at varying levels of geography. The Detailed DHC-B includes data for 300 detailed race and ethnicity groups and 1,187 American Indian and Alaska Native tribes and villages. Geographies include nation, state, county, census tract, place, and American Indian/Alaska Native/Native Hawaiian (AIANNH) areas. The Detailed DHC-A included total population and sex by age data for the same detailed races and ethnicities and geographies. The Census Bureau published the Detailed DHC-A in September 2023. In this section, we define relevant concepts, outline the statistics to be released, and then formulate the differentially private algorithm design problem.

2.1 Households

A household is an individual or group of individuals living together in an occupied housing unit. The Detailed DHC-B does not include vacant housing units. For our purposes, we treat “household” and “occupied housing unit” as interchangeable terms. In the CEF, every household has exactly one person designated as the householder. Every household has a household type (e.g., married couple family household) and a tenure status (e.g., renter occupied).

2.2 Geography

Every household is located in exactly one Census block that determines its geographic location. Census blocks are the most granular form of *geographic entities*. All other geographic entities (e.g., Los Angeles County, the state of California, and the United States) are aggregations of Census blocks. Geographic entities are divided into *geographic summary levels*. A geographic summary level is a set of nonoverlapping geographic entities, such as the set of all states or the set of all counties. The Detailed DHC-B produces statistics for the following geographic summary levels:

- Nation
- State or State equivalents
- County or County equivalents
- Census Tract

- Place
- AIANNH areas.

Henceforth, we tend to write State (County) without including the “or State (County) equivalent” qualifier, although one should assume the qualifier when applicable. Washington, D.C. is an example of a state equivalent. With some context-specific exceptions, this document adopts the convention of capitalizing references to levels and using lowercase when referencing an entity or entities within a level (e.g., there are over 3000 counties in the County level, and there are four counties in the state of Rhode Island).

2.3 Race and Ethnicity

In the 2020 Census, the Census Bureau collected detailed race and ethnicity data from individuals in accordance with the 1997 Federal Register Notice “Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity” released by the Office of Management and Budget (OMB).

Per these guidelines, every household is associated with one or more *race codes* of its householder and a single *ethnicity code* of its householder. That is, a household’s race and ethnicity assignment is based solely on the attributes of the householder, even if other individuals in the household have differing race or ethnicity codes. The maximum number of race codes, called the *race multiplicity*, that a householder, and hence household, can be associated with is limited to eight by the 2020 Census data collection procedures.

A *race group* is a set of race codes (e.g., German is defined by race codes ranging from 1170 through 1179). Similarly, an *ethnicity group* is a set of ethnicity codes (e.g., Mexican is defined by ethnicity codes ranging from 2010 through 2099). Hence, householders with differing race or ethnicity codes can nonetheless belong to the same race or ethnicity groups.

Detailed race or ethnicity groups are the most disaggregated racial or ethnic group classifications for which the Census Bureau publishes data. Examples of detailed racial or ethnic groups include Lebanese, Dutch, Guatemalan, Puerto Rican, Ethiopian, Nigerian, Mongolian, Thai, Brazilian, Belizean, Samoan, Marshallese, Chevak Native Village, and Navajo Nation. The *major* racial categories are aggregated race groupings that represent the minimum allowable racial categories for which census data may be published. The major racial categories in the 2020 Census were White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, and Some Other Race. The aggregated ethnic equivalent of the major racial categories is a coarse binary classification (Hispanic or Latino, Not Hispanic or Latino). *Regional* race or ethnicity groups provide an intermediate level of aggregation. Examples of regional racial or ethnic groups include European, Central American, Caribbean, Sub-Saharan African, Alaska Native, and American Indian. Subject-matter experts at the Census Bureau determined which race or ethnicity groups are detailed and which groups are regional. For the purposes of SafeTab-H, we take these classifications as given exogenous factors. The universe specification of valid detailed race and ethnicity groups and their classification into detailed or regional groups occurred before data collection for the 2020 Census [2]. Appendix G of [3] provides a complete enumeration of the detailed race and ethnicity groups.

A household is in a race group *alone* if all race codes associated with its householder are contained in the race group. For example, if a householder self-identifies with the single race code for Navajo Nation and no other race codes, the household would belong to the detailed race group Navajo Nation alone. Alternatively, a householder may report multiple race groups (e.g., British,

Scottish, and Dutch) that aggregate into the same regional group (European alone). A household is in a race group *alone or in any combination* if any race code associated with its householder is contained in the race group. This concept pertains to households where the householder self-identifies with a single detailed race (e.g., British) or with multiple detailed races (e.g., British and Thai). In both examples, the household belongs to the detailed British race group alone or in any combination. The household also belongs to the regional European race group alone or in any combination. Since all householders are only associated with a single ethnicity code, respondents may only be in one detailed ethnicity group and in one regional ethnicity group.

A *race characteristic iteration* is a race group combined with the specification of either “alone” or “alone or in any combination” (e.g., Latin American Indian alone or in any combination is a characteristic iteration). An *ethnicity characteristic iteration* is synonymous with an ethnicity group. Ethnicity characteristic iterations do not carry either “alone” or “alone or in any combination” designators. One household may be associated with multiple characteristic iterations. Like geographical entities, characteristic iterations are also divided into *characteristic iteration levels*. We have already provided the defining aspect of these iteration levels: namely, the concepts of *detailed* and *regional* race groups. We adopt the convention of capitalizing references to Detailed and Regional levels while using lowercase for references to detailed and regional iterations within a level. The Detailed characteristic iteration level consists of the set of characteristic iterations for all detailed race groups either alone or alone or in any combination (e.g., Japanese alone, Japanese alone or in any combination, Celtic alone, and Celtic alone or in any combination) and all detailed ethnicity groups. The Regional characteristic iteration level consists of the set of characteristic iterations for all regional race alone or alone or in any combination (e.g., Middle Eastern or North African alone, Middle Eastern or North African alone or in any combination, Polynesian alone, and Polynesian alone or in any combination) as well as all regional ethnicity groups. We intentionally omit the notion of a major race and ethnicity characteristic iteration level, as no statistics for this level are produced by the SafeTab-H algorithm for the Detailed DHC-B.

2.4 Population Groups

A *population group* is a pair (g, c) , where g is a geographic entity (e.g., the state of North Carolina), and c is a characteristic iteration (e.g., Latin American Indian alone or in any combination). Population groups are divided into *population group levels*. We will often identify a population group level by specifying a (geography level, characteristic iteration level) pair. However, each population group level is really a set of population groups, where each population group’s geographic entity belongs to the specified geography level and its characteristic iteration belongs to the specified characteristic iteration level. More formally, the Detailed DHC-B requires the publication of statistics for the following population group levels:

- (Nation, Detailed) $\equiv \{(g, c) : g \text{ is the nation, } c \text{ is a detailed characteristic iteration}\}$
- (State, Detailed) $\equiv \{(g, c) : g \text{ is a state, } c \text{ is a detailed characteristic iteration}\}$
- (County, Detailed) $\equiv \{(g, c) : g \text{ is a county, } c \text{ is a detailed characteristic iteration}\}$
- (Census Tract, Detailed) $\equiv \{(g, c) : g \text{ is a Census tract, } c \text{ is a detailed characteristic iteration}\}$
- (Place, Detailed) $\equiv \{(g, c) : g \text{ is a place, } c \text{ is a detailed characteristic iteration}\}$
- (AIANNH, Detailed) $\equiv \{(g, c) : g \text{ is an AIANNH area, } c \text{ is a detailed characteristic iteration}\}$

- (Nation, Regional) $\equiv \{(g, c) : g \text{ is the nation, } c \text{ is a regional characteristic iteration}\}$
- (State, Regional) $\equiv \{(g, c) : g \text{ is a state, } c \text{ is a regional characteristic iteration}\}$
- (County, Regional) $\equiv \{(g, c) : g \text{ is a county, } c \text{ is a regional characteristic iteration}\}$
- (Census Tract, Regional) $\equiv \{(g, c) : g \text{ is a Census tract, } c \text{ is a regional characteristic iteration}\}$
- (Place, Regional) $\equiv \{(g, c) : g \text{ is a place, } c \text{ is a regional characteristic iteration}\}$

In practice, some detailed or regional characteristic iterations may be omitted from the tabulations in a geography level. In other words, the above population group levels are supersets of the actual population group levels. This is done in accordance with specifications for the Detailed DHC-B provided by the Census Bureau. There is no concise representation for the exact level sets. The population group level (AIANNH, Regional) is intentionally omitted from the Detailed DHC-B.

One household may belong to multiple population groups in the set that comprises a population group level. For example, a household located in Texas with a householder reporting Kenyan and Ghanaian detailed races would belong to the (Texas, Kenyan alone or in any combination) and the (Texas, Ghanaian alone or in any combination) population groups which are both contained in the population group level identified by (State, Detailed). A household located in Schuyler County, NY with a householder reporting a single detailed race of Dutch would still belong to the (Schuyler County, NY, Dutch alone) and the (Schuyler County, NY, Dutch alone or in any combination) population groups which are both contained in the level identified by (County, Detailed). One household may be connected with population groups across multiple population group levels. The household with the Dutch householder residing in Schuyler County, NY would additionally belong to the (NY, Dutch alone) and (NY, Dutch alone or in any combination) population groups in the (State, Detailed) level, the (Schuyler County, NY, European alone) and (Schuyler County, NY, European alone or in any combination) population groups in the (County, Regional) level, etc. It is possible for a household to not belong to any population groups in a particular level. Specifically, because AIANNH areas do not cover the United States, a household located outside all designated AIANNH areas does not belong to any population groups in the (AIANNH, Detailed) level. For example, households in Arkansas do not belong to any AIANNH areas, and therefore would not contribute to any AIANNH area counts.

A householder's characteristic iterations primarily determine the number of population groups the household belongs to in each level because the geographic entities in a geography level are disjoint (a household cannot be located in both Schuyler County, NY and Fairfax County, VA). A household associated with the maximum of eight race codes and a single ethnicity code could belong to at most nine detailed characteristic iterations (eight alone or in any combination race groups and one ethnic group) and, similarly, at most nine regional characteristic iterations. Thus, for any given population group level, the maximum number of population groups a household may contribute to is nine.

2.5 Detailed Demographic and Housing Characteristics File B

The Detailed DHC-B aims to tabulate statistics about household type and tenure by population groups. The following statistical tables are released for each eligible population group.¹

¹Population groups may be deemed ineligible to receive certain statistics for various reasons discussed throughout the document. For instance, Detailed DHC-B tables are not released for groups that do not appear in the Detailed DHC-A's T01001.

- Household type counts for eligible population groups. The household type tables come in four different variants (T03001, T03002, T03003, and T03004) as shown in Tables 1, 2, 3, and 4. For convenience in mathematical expressions, we use HT to refer generically to the household type table class.
- Tenure counts for eligible population groups. The tenure tables come in two different variants (T04001 and T04002) as shown in Tables 5 and 6. For convenience in mathematical expressions, we use the abbreviation T to refer generically to the tenure table class.

Each table has a *basis*, a set of fine-grained, disaggregated table cells from which the remaining cells may be generated. In the below table shells, the dark text indicates which cells form the table's basis, while the light text shows the aggregated table cells. For a given table, every household can be assigned to exactly one category of the table's basis. For example, with T03002, every household is exclusively either a family household or a nonfamily household.

T03001. Household Type (Universe)

Universe: Households

Total

Table 1: T03001 consists solely of a total household count.

T03002. Household Type (2 Categories)

Universe: Households

Total:

Family Household

Nonfamily Household

Table 2: T03002 consists of counts for family household and nonfamily household.

T03003. Household Type (6 Categories)

Universe: Households

Total:

Family Household:

Married Couple Family

Other Family

Nonfamily Household:

Householder Living Alone

Householder Not Living Alone

Table 3: T03003 contains counts for married couple households, other family households, and the breakdown of nonfamily households into categories for householders living alone or not living alone.

T03004. Household Type (8 Categories)

Universe: Households

Total:

Family Household:

Married Couple Family

Other Family:

Male householder, no spouse present

Female householder, no spouse present

Nonfamily Household:

Householder Living Alone

Householder Not Living Alone

Table 4: T03004 contains counts for married couple households, the breakdown of other family households into categories for male or female householders with no spouse present, and the breakdown of nonfamily households into categories for householders living alone or not living alone.

T04001. Tenure (Universe)

Universe: Occupied Housing Units

Total

Table 5: T04001 consists solely of a total household count.

T04002. Tenure (3 Categories)

Universe: Occupied Housing Units

Total:

Owned with a mortgage or a loan

Owned free and clear

Renter Occupied

Table 6: T04002 consists of counts for three household tenure categories: owned with a mortgage or a loan, owned free and clear, and renter occupied.

As previously mentioned, despite the difference in universe terminology (i.e., household vs occupied housing units), the household type tables and tenure tables both provide total counts over the same entities. That is, a population group’s total household count (e.g., T03001) measures the same quantity as the population group’s total occupied housing unit count (e.g., T04001).

An additional table pertaining to properties of persons is also needed as an input for the SafeTab-H algorithm. The T01001 table is produced as part of the Detailed DHC-A.

T01001 Total Population

Universe: Total Population

Total

Table 7: The T01001 table contains counts of persons, instead of households, and is iterated by detailed race and ethnicity of each individual in the population rather than of the householder. This table is an output of SafeTab-P and was released in the Detailed DHC-A.

2.6 Private Release Problem

The Census Bureau release of statistical data products is regulated under Title 13, which disallows any data publications in which an individual’s data can be identified [4]. Moreover, it has been demonstrated that legacy statistical disclosure limitation techniques are vulnerable to attacks that can reconstruct the sensitive person records from aggregate statistics [5]. Hence, the Census Bureau decided to release many of the 2020 Census data products, including the Detailed DHC-B, using algorithms that satisfy modern privacy definitions like differential privacy [6].

In particular, we describe a disclosure avoidance technique for the Detailed DHC-B that was designed to satisfy the following desiderata:

- *Privacy*: The algorithm must satisfy a variant of differential privacy known as zero-concentrated differential privacy (zCDP) with respect to arbitrary changes of any household record’s values.
- *Population Groups*: The algorithm must release statistics for a predefined set of race and ethnicity characteristic iterations and the following geography levels: Nation, State, County, Census Tract, Place, and AIANNH areas. The algorithm also produces statistics for equivalent geographic regions in Puerto Rico.
- *Adaptivity*: For each population group, the algorithm may adaptively choose which granularity of the household type and tenure tables to release. The selection of granularity depends on the T01001 total population counts of each population group. For instance, a population group with a few people may only receive a total household count (i.e., T03001 and T04001 tables), while a population group with many people may receive the full table shells (i.e., T03004 and T04002 tables).
- *Accuracy*: The algorithm must achieve pre-specified accuracy levels for population groups in terms of the margins of error (MOE) in output counts. Different population groups may have different MOEs specified (described later in the paper in Table 9). The MOE discussed in the paper captures error induced by disclosure avoidance alone and does not capture other sources of error such as under counts in the 2020 Census.
- *Integrity*: The output statistics must be integers.
- *Minimal Consistency*: The algorithm is not required, in general, to ensure consistency. That is, different counts output by the system need not be consistent with each other (e.g., the number of households of a certain characteristic iteration in the United States need not equal the sum of the household counts for the same characteristic iteration across all states). We also note that no consistency is enforced with other data products, such as the DHC. However, some postprocessing of outputs is done to address specific demographic reasonableness concerns. These postprocessing steps are discussed in Section 6.3.

In the rest of the paper, we describe the SafeTab-H differential privacy algorithm, discuss implementation and parameter tuning, and analyze bounds on the privacy loss achievable while satisfying the constraints mentioned above.

3 SafeTab-H Algorithm

SafeTab-H is a privacy algorithm for releasing 2020 Census household type and tenure counts, iterated by detailed race and ethnicity characteristic iterations. This section covers a simplified

abstraction of the SafeTab-H algorithm. Additional implementation details are discussed in Section 6. In this section, we describe the algorithm as applied to the United States. Puerto Rico is discussed in Section 6.5.2. The algorithm acts on a private dataframe of household records derived from the 2020 Census. At a high level, the algorithm performs the following steps for each eligible population group:

1. It selects a household type table variant (T03001, T03002, T03003, or T03004) and a tenure table variant (T04001 or T04002).
2. It queries the household data to compute the basis of each selected table variant.
3. It perturbs the computed basis of each selected table variant by adding noise drawn from a discrete Gaussian distribution.

The selection in the first step is fully determined by a comparison of each eligible population group’s T01001 total population count to population thresholds pre-specified by the Census Bureau. The computation in the second step produces accurate counts according to the 2020 CEF. Finally, the noise infusion in the third step provides necessary privacy protection in accordance with Census Bureau policy.

For a more technical look at SafeTab-H, we begin with a description of the algorithm’s input data sources and then present a pseudocode representation.

3.1 Input Data Description

3.1.1 Household Data

Household records are stored in the 2020 CEF, a relational database with multiple person and household attributes spread across several linked dataframes. Many of these attributes are irrelevant to the tabulations in the Detailed DHC-B. As such, we assume a simplified, reduced-form data representation that is sufficient for our purposes. This dataset is a private dataframe derived from the 2020 CEF that consists of a row for each household in the United States with the following attributes: BlockID, RaceEth, HouseholdType, and Tenure.

BlockID is a single attribute that geolocates a household record to a unique Census block. As previously discussed, all geographic entities are aggregations of blocks. Thus, we assume that BlockID implicitly encodes each record’s unique Census tract, county, and state. BlockID also encodes whether a record belongs to a place or AIANNH area and, if so, uniquely identifies its place or AIANNH area. We note that all records are vacuously included in the nation geographic entity.

RaceEth is a single attribute that encodes up to eight race codes and an ethnicity code of the householder. That is, one household’s RaceEth attribute may indicate the householder is Andorran and Dominican while another household’s RaceEth attribute indicates a householder that is South African, Japanese, Tongan, and Not Hispanic or Latino. We assume this RaceEth conceptualization combined with Census Bureau specifications fully determines the characteristic iterations of a household.

HouseholdType categorizes the relationship of the householder to other members of their household. There are five categories:

- Married couple family.
- Other family, male householder with no spouse present.
- Other family, female householder with no spouse present.
- Nonfamily with the householder living alone.
- Nonfamily with the householder not living alone.

Tenure categorizes the householder’s owner or renter status. There are three categories:

- Owned with a mortgage or a loan.
- Owned free and clear.
- Renter occupied.

3.1.2 Total Population Counts

The Detailed DHC-A’s T01001 table contains total population counts by detailed race and ethnicity. These counts are used by SafeTab-H to determine the granularity of household type and tenure tabulations each population group should receive. Since the Detailed DHC-A was published before the Detailed DHC-B, for the purposes of SafeTab-H, we treat the T01001 counts as fixed exogenous inputs to the program. As with the household data, we assume a reduced-form dataframe representation. In this case, the dataframe consists of a row for each population group output by the SafeTab-P privacy algorithm. This dataset has the following attributes: PopGroup and Count.

PopGroup is a single attribute that encodes the geographic entity and the characteristic iteration of a population group.

Count is the population group’s total population count from the Detailed DHC-A.

Equivalently, we can view the T01001 input data as a function that maps population groups to their corresponding total population counts.

3.2 The Algorithm Description

For reference, the notation used in this section and the algorithm pseudocode is summarized in Table 8.

SafeTab-H produces tabulations for population groups. Population groups are split into sets called population group levels (specified by a geography level and an iteration level) with distinct privacy-loss budgets for each table class (household type and tenure). Records are associated with population groups via transformations that map their BlockID to geographic entities, and their RaceEth attribute to characteristic iterations. For the purposes of this section, we assume the following model for population groups:

- SafeTab-H is given fixed T01001 population group counts. That is, there is a set of population groups \mathcal{T} and a mapping $h : \mathcal{T} \rightarrow \mathbb{Z}$ of population groups to fixed counts of their total population.
- SafeTab-H operates on population group levels $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_\omega$. For example, \mathcal{P}_i may be the level (State, Detailed) consisting of population groups, such as (Iowa, Albanian alone) and (Kansas, German alone or in any combination).
- SafeTab-H accounts for the possibility that not all population groups in the input specification have corresponding T01001 counts. Population groups without T01001 counts do not receive household type or tenure results in the Detailed DHC-B. That is, SafeTab-H produces household type and tenure tables for each population group $P \in \mathcal{P}_i \cap \mathcal{T}$ for $1 \leq i \leq \omega$.
- In SafeTab-H, the household type (HT) and tenure (T) counts receive separate privacy-loss budgets for each population group level $\rho_1^t, \rho_2^t, \dots, \rho_\omega^t$ with ρ_i^t corresponding to the budget for population group level \mathcal{P}_i for table class $t \in \{\text{HT}, \text{T}\}$. Privacy-loss budgets are described in greater detail in Section 4. For now, we note that each ρ_i^t is a positive real-valued number.
- For each \mathcal{P}_i , we assume we have a function $g_i : \mathcal{I} \rightarrow 2^{\mathcal{P}_i}$, where \mathcal{I} is the domain of household records in the private dataframe. That is, g_i maps a household record r to the subset of population groups at the level i to which it belongs (i.e., $g_i(r) \subset \mathcal{P}_i$). For example, suppose i corresponds to the (State, Regional) level, the record r 's BlockID uniquely identifies its state as Idaho, and its RaceEth attribute encodes Nigerian, Beninese, Tongan, and Not Hispanic or Latino. Then $g_i(r)$ would associate the record with the following population groups: (Idaho, Sub-Saharan African alone or in any combination) and (Idaho, Polynesian alone or in any combination).
- We assume the stability of g_i , denoted by $\Delta(g_i)$, is known. The stability is defined as the maximum number of population groups a record could belong to in a level. Formally, $\Delta(g_i) = \max_{r \in \mathcal{I}} |g_i(r)|$ [7]. Importantly, this value defines what could be the maximum based on any hypothetically possible record, rather than defining what is the maximum based on the collected 2020 Census data. In other words, the stability is a data-independent value. As described earlier, this value is $\Delta(g_i) = 9$ for all population group levels tabulated in the Detailed DHC-B.

The main algorithm is presented in Algorithm 1. This algorithm proceeds by looping over the population group levels. For each population group level, we apply g_i to the dataframe to map each record to the set of population groups it is associated with. Then, for each population group in the level, we call the function `VECTORIZEPOPULATIONGROUP`, passing in a dataframe containing just the records in that population group. However, we skip over population groups that do not have T01001 counts.

The pseudocode for the procedure `VECTORIZEPOPULATIONGROUP` is given in Algorithm 2. This code performs the high-level steps 1 and 2 discussed earlier, returning two basis vectors (one for a household type table and another for a tenure table) for the population group. It starts by comparing the provided T01001 count against a set of given thresholds, denoted θ_1, θ_2 , and θ_3 for the HT table class and ψ_1 for the T table class. Depending on which thresholds the published count exceeds, it selects an HT table variant (T03001, T03002, T03003, or T03004) and a T table variant (T04001 or T04002). It then computes the basis vectors for the selected household type and tenure tables by counting how many households in the population group belong to each component of the respective basis vectors.

Notation	Description
ω	the number of population group levels
\mathcal{P}_i	population group level i
\mathcal{T}	population groups with T1 counts
h	a mapping of population groups to their T1 counts
ρ_i^t	the privacy-loss budget allocated to population group level i for table $t \in \{\text{HT}, \text{T}\}$
g_i	a function mapping records to the set of population groups in \mathcal{P}_i to which the record belongs
$\Delta(g_i)$	$\max_{r \in \mathcal{I}} g_i(r) $
$[a, b]$	The range of integers starting with a and ending with b , inclusive.

Table 8: A summary of the notation used in Section 3

The computed HT vectors of each population group in the population group level are stacked to create one large HT vector for the population group level. Similarly, one large T vector is constructed.

The HT and T vectors for the population group level are then passed through the VECTORDISCRETEGAUSSIAN procedure using the level’s privacy-loss budget for the respective table classes.

The pseudocode for the procedure VECTORDISCRETEGAUSSIAN is given in Algorithm 3. This procedure carries out the high-level step 3 by adding independently drawn noise from the discrete Gaussian distribution to each of the input vector’s components. The distribution is scaled according to the privacy-loss budget input to the procedure.

Next, we introduce background material on zCDP so that we can analyze the privacy guarantees of the SafeTab-H algorithm.

Algorithm 1 The main SafeTab-H algorithm.

Input: df : A private dataframe with attributes [BlockID, RaceEth, HouseholdType, Tenure] and one row for each household in the United States.

Input: $\{\rho_i^t\}_{i \in [1, \omega]}$: Privacy-loss parameters for each population-group level $i \in [1, \omega]$ and table class $t \in \{HT, T\}$.

Input: (\mathcal{T}, h) : A mapping h of population groups to their T01001 total population counts, along with the mapping's domain \mathcal{T} .

```
1: procedure SAFETAB-H( $df, \{\rho_i^t\}, \mathcal{T}, h$ )
2:   for  $i \in [1, \omega]$  do
3:      $df_i \leftarrow df.flatmap(g_i)$ ;            $\triangleright df_i$  has schema [PopGroup, HouseholdType, Tenure]
4:      $s \leftarrow \Delta(g_i)$                         $\triangleright 1$  row in  $df$  may result in  $\leq s$  rows in  $df_i$ 
5:      $V_{HT}, V_T \leftarrow []$                     $\triangleright$  Initialize empty vectors of counts
6:     for  $P \in \mathcal{P}_i$  do
7:       if  $P \notin \mathcal{T}$  then
8:         continue                                $\triangleright$  Skip population groups that do not have T01001 counts
9:       end if
10:       $c \leftarrow h(P)$                             $\triangleright$  Get the T01001 count of population group  $P$ 
11:       $df_P \leftarrow df_i.filter(PopGroup == P)$ 
12:       $v_{HT}, v_T \leftarrow \text{VECTORIZEPOPULATIONGROUP}(df_P, P, c)$ 
13:       $V_{HT}.append(v_{HT})$ 
14:       $V_T.append(v_T)$ 
15:    end for
16:    Output  $\text{VECTORDISCRETEGAUSSIAN}(V_{HT}, \rho_i^{HT} / s)$ 
17:    Output  $\text{VECTORDISCRETEGAUSSIAN}(V_T, \rho_i^T / s)$ 
18:  end for
19: end procedure
```

Algorithm 2 Subroutine of SafeTab-H that returns a vector of HT counts and a vector of T counts for a single population group.

Input: df : A private dataframe with attributes [PopGroup, HouseholdType, Tenure]. This dataframe should only contain household records in population group P .

Input: P : The population group.

Input: c : The T01001 count of population group P .

```

1: procedure VECTORIZEPOPULATIONGROUP( $df, P, c$ )
2:   // Adaptive process for Household Type tabulations
3:    $v_{HT}, v_T \leftarrow []$ 
4:   if  $c > \theta_3$  then
5:     // Vectorize basis of T03004
6:      $v \leftarrow df.map(\text{HouseholdType} \rightarrow \text{T03004 basis}).groupby(\text{T03004 basis}).count()$ 
7:      $v_{HT}.append(v)$ 
8:   else if  $c > \theta_2$  then
9:     // Vectorize basis of T03003
10:     $v \leftarrow df.map(\text{HouseholdType} \rightarrow \text{T03003 basis}).groupby(\text{T03003 basis}).count()$ 
11:     $v_{HT}.append(v)$ 
12:  else if  $c > \theta_1$  then
13:    // Vectorize basis of T03002
14:     $v \leftarrow df.map(\text{HouseholdType} \rightarrow \text{T03002 basis}).groupby(\text{T03002 basis}).count()$ 
15:     $v_{HT}.append(v)$ 
16:  else
17:    // Vectorize basis of T03001
18:     $v_{HT}.append(df.count())$ 
19:  end if
20:  // Adaptive process for Tenure tabulations
21:  if  $c > \psi_1$  then
22:    // Vectorize basis of T04002
23:     $v \leftarrow df.map(\text{Tenure} \rightarrow \text{T04002 basis}).groupby(\text{T04002 basis}).count()$ 
24:     $v_T.append(v)$ 
25:  else
26:    // Vectorize basis of T04001
27:     $v_T.append(df.count())$ 
28:  end if
29:  Return  $v_{HT}, v_T$ 
30: end procedure

```

Algorithm 3 The discrete Gaussian mechanism for vectors.

Input: a : An n dimensional vector of integers.

Input: ρ : A privacy-loss parameter.

```

1: procedure VECTORDISCRETEGAUSSIAN( $a, \rho$ )
2:    $y \leftarrow \mathcal{N}_{\mathbb{Z}}^n \left( \frac{1}{2\rho} \right)$ 
3:   return  $a + y$ 
4: end procedure

```

4 Privacy Preliminaries

In this section, we give necessary background on zCDP and the privacy properties that it guarantees.

4.1 Privacy Definitions

Definition 1 (Neighboring Databases). Let x, x' be databases represented as multisets of tuples. We say that x and x' are *neighbors* if their symmetric difference is 1.

Definition 2 (Bounded-Neighboring Databases). Let x, x' be databases represented as multisets of tuples. We say that x and x' are *bounded neighbors* if they differ by *arbitrarily changing* at most one tuple.

We sometimes refer to neighboring databases as “unbounded neighbors” to differentiate between Definitions 1 and 2. However, it should be assumed we are referring to Definition 1 unless a distinction is explicitly made.

The mathematical definition of zCDP expresses a bound on the *Rényi divergence* between the distributions of a mechanism run on neighboring databases.

Definition 3. The *Rényi divergence of order α* between distribution P and distribution Q , denoted $D_\alpha(P\|Q)$, is defined as

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \left(\mathbb{E}_{x \sim P} \left[\left(\frac{P(x)}{Q(x)} \right)^{\alpha-1} \right] \right) \quad (1)$$

Definition 4. (zCDP [8]) An algorithm $M : \mathcal{X} \rightarrow \mathcal{Y}$ satisfies ρ -zCDP if for all neighboring $x, x' \in \mathcal{X}$ and for all $\alpha \in (1, \infty)$,

$$D_\alpha(M(x)\|M(x')) \leq \rho\alpha. \quad (2)$$

We also define bounded zCDP, which considers bounded-neighboring databases instead of unbounded-neighboring databases.

Definition 5. (Bounded zCDP [8]) An algorithm $M : \mathcal{X} \rightarrow \mathcal{Y}$ satisfies ρ -zCDP if for all bounded neighbors $x, x' \in \mathcal{X}$ and for all $\alpha \in (1, \infty)$,

$$D_\alpha(M(x)\|M(x')) \leq \rho\alpha. \quad (3)$$

4.2 Privacy Properties

4.2.1 Composition

One of the most useful and important properties of privacy definitions is their behavior under composition. For the SafeTab-H privacy analysis, sequential composition for zCDP is sufficient.

Lemma 1. (*Adaptive sequential composition of zCDP [8]*) Let $M_1 : \mathcal{X} \rightarrow \mathcal{Y}$ and $M_2 : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ be mechanisms satisfying ρ_1 -zCDP and ρ_2 -zCDP respectively. Let $M_3(x) = M_2(x, M_1(x))$. Then M_3 satisfies $(\rho_1 + \rho_2)$ -zCDP.

4.2.2 Postprocessing

Zero-concentrated differential privacy is closed under postprocessing, meaning that the privacy guarantee cannot be weakened by manipulating the outputs of a zCDP mechanism without reference to the protected inputs.

Lemma 2. (Postprocessing for zCDP [8]) Let $M : \mathcal{X} \rightarrow \mathcal{Y}$ and $f : \mathcal{Y} \rightarrow \mathcal{Z}$ be randomized algorithms. Suppose M satisfies ρ -zCDP. Then $f \circ M : \mathcal{X} \rightarrow \mathcal{Z}$ satisfies ρ -zCDP.

4.3 Base Mechanism

Definition 6 (L2 Sensitivity). Given a vector function $q : \mathcal{X} \rightarrow \mathbb{Z}^n$, the sensitivity of q is $\sup_{x, x'} \|q(x) - q(x')\|_2$ where x and x' are neighboring databases and $\|\cdot\|_2$ is the Euclidean norm.

There is an equivalent notion of bounded sensitivity.

Definition 7 (Bounded L2 Sensitivity). Given a vector function $q : \mathcal{X} \rightarrow \mathbb{Z}^n$, the sensitivity of q is $\sup_{x, x'} \|q(x) - q(x')\|_2$ where x and x' are bounded-neighboring databases and $\|\cdot\|_2$ is the Euclidean norm.

Definition 8. The discrete Gaussian distribution $\mathcal{N}_{\mathbb{Z}}(\sigma^2)$ centered at 0 is

$$\forall x \in \mathbb{Z}, \quad \Pr[X = x] = \frac{e^{-x^2/2\sigma^2}}{\sum_{y \in \mathbb{Z}} e^{-y^2/2\sigma^2}}. \quad (4)$$

Lemma 3. [9] Let $q : \mathcal{X} \rightarrow \mathbb{Z}^n$ with sensitivity Δ . Then outputting `VECTORDISCRETEGAUSSIAN`($q(x), \rho$) from Algorithm 3 satisfies $\Delta^2 \rho$ -zCDP.

5 SafeTab-H Privacy Analysis

In this section, we show that the SafeTab-H algorithm presented in Section 3 satisfies zero-concentrated differential privacy (zCDP). We then demonstrate that SafeTab-H also satisfies bounded zCDP with a privacy loss that increases by a factor of 2.

Theorem 1. Let $\rho_{total} = \sum_{i=1}^{\omega} \rho_i^{HT} + \rho_i^T$. Algorithm 1 satisfies ρ_{total} -zCDP with respect to its inputs.

Proof. The proof of Theorem 1 follows via the combination of sensitivity analysis along with the fact that the base mechanism, `VECTORDISCRETEGAUSSIAN`, satisfies zCDP.

We first argue the transformation `VECTORIZEPOPULATIONGROUP` in Algorithm 2 has sensitivity 1 with respect to each of the outputs v_{HT}, v_T .

We decompose our argument into two subclaims.

Subclaim 1: The vector, v_{HT} , produced in Lines 2-19 of `VECTORIZEPOPULATIONGROUP` can have at most one of its elements increased by 1 due to the addition of a record or decreased by 1 due to the removal of a record.

Proof of subclaim 1: Assume the population group P and its corresponding T01001 count c are fixed.

If $c > \theta_1$, there are 3 similar cases. We show the case where $c > \theta_3$ and omit the others.

In `VECTORIZEPOPULATIONGROUP`, each record in the input dataframe for population group P maps to exactly one group of the T03004 basis groupby. Since v_{HT} is a vector of the counts of each group, it follows that the addition of a single record can increase one element of v_{HT} by at

most 1 or, alternatively, the removal of a single record can decrease one element of v_{HT} by at most 1.

If $c \leq \theta_1$, the vector v_{HT} is a size 1 vector containing only the total count. It follows that the addition of a single record can increase the one element of v_{HT} by at most 1. Alternatively, the removal of a single record can decrease the one element of v_{HT} by at most 1.

This proves subclaim 1.

Subclaim 2: The vector, v_T , produced in Lines 20-28 of VECTORIZEPOPULATIONGROUP can have at most one of its elements increased by 1 due to the addition of a record or decreased by 1 due to the removal of a record. We omit the proof, as it is nearly identical to that of subclaim 1.

Next, we claim that the i th loop of the **for** loop on line 2 of Algorithm 1 satisfies $(\rho_i^{HT} + \rho_i^T)$ -zCDP. By the definition of s , any particular record can appear in the input (df_P) in at most s calls to VECTORIZEPOPULATIONGROUP. The vectors V_{HT}, V_T are the combination of each of the vectors produced by VECTORIZEPOPULATIONGROUP. Since a single record can appear in s calls to VECTORIZEPOPULATIONGROUP and each call can change a single element by at most 1, the addition or removal of a single record can change at most s elements in V_{HT}, V_T by at most 1. Therefore $\|V_{HT} - V'_{HT}\|_2 \leq \sqrt{s}$, where V_{HT} and V'_{HT} are data vectors resulting from neighboring databases. The same holds for the vector V_T . Thus, by Lemma 3, the call to VECTORDISCRETEGAUSSIAN with privacy-loss parameter ρ_i^{HT}/s on Line 16 satisfies ρ_i^{HT} -zCDP. Similarly, the call on Line 17 with privacy-loss parameter ρ_i^T/s satisfies ρ_i^T -zCDP. By sequential composition (Lemma 1), the combination satisfies $(\rho_i^{HT} + \rho_i^T)$ -zCDP.

Finally, the overall algorithm satisfies $(\sum_{i=1}^{\omega} \rho_i^{HT} + \rho_i^T)$ -zCDP by Lemma 1. \square

5.1 Converting to Bounded zCDP

Theorem 2. Let $\rho_{total} = \sum_{i=1}^{\omega} \rho_i^{HT} + \rho_i^T$. Algorithm 1 satisfies bounded $2\rho_{total}$ -zCDP with respect to its inputs.

Proof. The proof of Theorem 2 follows via the combination of sensitivity analysis along with the fact that the base mechanism, VECTORDISCRETEGAUSSIAN, satisfies zCDP. Under bounded zCDP, it is sufficient to bound the changes due to the removal of one record *and* addition of another. As a result, we reason specifically about the effects of adding and removing a record separately and the effects when both happen.

We first argue the transformation VECTORIZEPOPULATIONGROUP in Algorithm 2 has sensitivity 1 with respect to each of the outputs v_{HT}, v_T .

We decompose our argument into two subclaims. Namely,

Subclaim 1: The vector, v_{HT} , produced in Lines 2-19 of VECTORIZEPOPULATIONGROUP can have at most one of its elements increased by 1 due to the addition of a record or decreased by 1 due to the removal of a record.

Proof of subclaim 1: Assume the population group P and its corresponding T01001 count c are fixed. If $c > \theta_1$, there are 3 similar cases. We show the case where $c > \theta_3$ and omit the others.

In VECTORIZEPOPULATIONGROUP, each record in the input dataframe for population group P maps to exactly one group of the T03004 basis groupby. Since v_{HT} is a vector of the counts of each group, it follows that the addition of a single record can increase one element of v_{HT} by at most 1 or, alternatively, the removal of a single record can decrease one element of v_{HT} by at most 1.

If $c \leq \theta_1$, the vector v_{HT} is a size 1 vector containing only the total count. It follows that the addition of a single record can increase the one element of v_{HT} by at most 1. Alternatively, the removal of a single record can decrease the one element of v_{HT} by at most 1.

This proves subclaim 1.

Subclaim 2: The vector, v_T , produced in Lines 20-28 of VECTORIZEPOPULATIONGROUP can have at most one of its elements increased by 1 due to the addition of a record or decreased by 1 due to the removal of a record. We omit the proof, as it is nearly identical to that of Subclaim 1.

Next, we claim that the i th loop of the **for** loop on line 2 of Algorithm 1 satisfies bounded $(2\rho_i^{HT} + 2\rho_i^T)$ -zCDP. By the definition of s , any particular record can appear in the input (df_P) in at most s calls to VECTORIZEPOPULATIONGROUP. The vectors V_{HT}, V_T are the combination of each of the vectors produced by VECTORIZEPOPULATIONGROUP. Since a single record can appear in s calls to VECTORIZEPOPULATIONGROUP and each call can change a single element by at most 1, the addition or removal of a single record can change at most s elements in V_{HT}, V_T by at most 1. Since the addition of one record can increase at most s elements by 1 and the removal of one record can decrease at most s elements by 1, we have $\|V_{HT} - V'_{HT}\|_2 \leq \sqrt{2s}$, where V_{HT} and V'_{HT} are data vectors resulting from bounded-neighboring databases. The same holds for the vector V_T . Thus, by Lemma 3, the call to VECTORDISCRETEGAUSSIAN with privacy-loss parameter ρ_i^{HT}/s , on Line 16 satisfies bounded $2\rho_i^{HT}$ -zCDP. Similarly, the call on Line 17 with privacy-loss parameter ρ_i^T/s satisfies bounded $2\rho_i^T$ -zCDP. By sequential composition (Lemma 1), the combination satisfies bounded $(2\rho_i^{HT} + 2\rho_i^T)$ -zCDP.

Finally, the overall algorithm satisfies bounded $(\sum_{i=1}^{\omega} 2\rho_i^{HT} + 2\rho_i^T)$ -zCDP by Lemma 1. \square

6 Implementation of SafeTab-H

The algorithm presented in Section 3 is a simplified version of the implemented SafeTab-H program. In this section, we describe some of the differences between the implementation and the simplified algorithm. We focus on differences that could affect the privacy calculus and describe why the implementation is equivalent to the simplified algorithm.

6.1 Input Validation

Input validation is an important step before deploying a differentially private algorithm. SafeTab-H performs extensive validation of its input data to ensure that provided data are in the expected formats and are internally consistent. There is not an impact to any privacy guarantees when input validation is done for the public datasets, like the list of all geographic entities for which data is to be tabulated. The input validation on private datasets that contain the data of individual census respondents has privacy implications because validation failures can reveal information about the dataset. However, validation failures are only visible to the trusted curator running the program, and on failure, no part of the differentially private program is run. Validation failures are made available to the trusted curator, so they can correct any errors in the provided input files before executing the differentially private program. Failures in validation are not released publicly and therefore, do not contribute to any privacy loss.

6.2 Tumult Analytics

Rather than directly calculating stability and sampling from noise distributions, SafeTab-H is implemented using Tumult Analytics[10], a framework for implementing differentially private queries.

A key benefit of using Tumult Analytics is that all access to the sensitive data is mediated through a Tumult Analytics `Session`. The `Session` tracks all the transformations and measure-

ments performed on the sensitive data and is able to correctly compute the total privacy loss of the computation on the sensitive data. In SafeTab-H, we construct an `Analytics Session` with:

- The total privacy-loss budget for the pipeline (calculated as the sum of all budgets ρ_i).
- The private dataset of household records.
- The public datasets (information on all detailed race and ethnicity groups, characteristic iterations, geographic entities, and total population counts for each population group).
- The neighboring definition (privacy is with respect to the addition/removal of one record from the private dataset). Note that Tumult Analytics does not support bounded neighbors.
- The privacy definition to be satisfied (e.g., zCDP).

We then implement all data transformations (like mapping a record to its characteristic iterations) and queries within the framework. Tumult Analytics tracks the stability throughout the transformations and applies an appropriate amount of noise to the final queries, guaranteeing that the outputs are differentially private and no more than the total budget is expended.

The use of Tumult Analytics allows (and necessitates) some other deviations from the simplified algorithm. Rather than vectorizing each population group sequentially in a for-loop, we use Analytics' `groupby` feature to tabulate many population groups at once. Analytics requires users of the `groupby` feature to specify all groups to tabulate in advance in the form of a `KeySet` object. In the case of SafeTab-H, we construct `KeySets` containing the combinations of possible geographic areas, characteristic iterations, and either household types or tenure categories. We build these `KeySets` using the T01001 total population counts output by SafeTab-P (rather than relying on observed groups present in the 2020 CEF). `KeySets` do not use the confidential 2020 CEF data to ensure that we do not reveal whether population groups are empty via their presence or absence in the output data. We note the simplified algorithm also allowed for the possibility that df_P (the filtered dataframe containing records associated with population group P) is empty.

6.3 Postprocessing for Addressing Demographic Reasonableness Concerns

After the differentially private algorithm has completed, we perform several additional postprocessing steps. Because these steps are purely postprocessing, they cannot affect the differential privacy guarantees per Lemma 2. These postprocessing steps are designed to address specific data quality concerns that arise when adding noise to tabular statistics. All postprocessing was implemented under the direction of subject-matter experts at the Census Bureau. These steps do not exhaustively address all possible demographic reasonableness concerns.

6.3.1 Marginals

As mentioned in Section 3, SafeTab-H selects different table variants for each population group based on their T01001 counts and then produces noisy estimates of the selected table's basis. For example, if T03003 were selected for a population group, the pseudocode would produce estimates for the categories "Married Couple Family", "Other Family", "Nonfamily: Householder Living Alone", and "Nonfamily: Householder Not Living Alone" but would not produce values for "Family Household", "Nonfamily Household", and "Total". To ensure that the selected table variant is released, SafeTab-H has a postprocessing step that aggregates each table basis to construct the complete table shell. In our example, "Married Couple Family" and "Other Family" are

summed to produce a noisy estimate for the "Family Household" table cell, "Nonfamily: Householder Living Alone" and "Nonfamily: Householder Not Living Alone" are summed to produce a noisy estimate for the "Nonfamily Household" table cell, and all four basis values are summed to produce a noisy estimate for the "Total" table cell. This ensures the complete T03003 table shell is output by SafeTab-H and that the table marginals for a population group are consistent with the table's basis.

6.3.2 Suppression

SafeTab-H does not implement its own suppression. However, SafeTab-H does not produce statistics for any population group suppressed by SafeTab-P (because SafeTab-H relies on the population group totals that have undergone suppression). Thus, some population groups that might otherwise be expected to appear in the output are missing. These suppressed counts are not published in any fashion in the S-DHC.

Suppressing outputs based on the noisy counts produced by SafeTab-P (as opposed to the noisy counts calculated by SafeTab-H) does not introduce bias into the counts that are directly published within the Detailed DHC-B. Additionally, the "suppressed" population groups are completely deterministic with respect to a fixed SafeTab-P output, so there is no randomness in the process.

6.3.3 Coterminous Geographies

Sometimes, two or more geographic entities in different geographic summary levels share the same geographic boundaries (i.e., they are aggregated from identical collections of Census blocks). For example, Washington, D.C. is tabulated as a state, county, and place. These geographic entities are called *coterminous*. Another coterminous example is a county containing a single Census tract. A characteristic iteration receives different independent noisy measurements for each geographic summary level of a given coterminous area. However, its counts should be identical at each summary level. Some geographic entities at different summary levels that do not share the same geographic boundaries should still be statistically equivalent. For example, if a county contains one water-only tract and one nonwater tract, characteristic iterations should have the same counts in the nonwater tract as in the county. We also consider statistically equivalent geographic areas to be coterminous. Subject-matter experts at the Census Bureau created a postprocessing step (implemented as a standalone program) that corrects inconsistencies in coterminous geographic entities. Since this correction is performed outside the DAS, it is out of scope for this paper, but we note that as pure postprocessing, it does not impact the privacy analysis.

6.3.4 Tabulation System Suppression

Additional demographic reasonableness corrections are addressed outside the DAS. In particular, the Decennial Tabulation System also performs suppression postprocessing. Again, per Lemma 2, this does not impact the privacy analysis. The suppression conducted by non-DAS systems is out of scope for this paper but includes further enforcement of nonnegativity as well as suppression in cases where noisy counts for alone characteristic iterations appear to be greater than the corresponding noisy counts for alone or in any combination characteristic iterations. These suppressed counts are published as "X" in the S-DHC.

6.4 Input Sourcing

Section 3 describes the total population counts as being sourced by the Detailed DHC-A. In reality, this data is sourced by the outputs of the SafeTab-P privacy program that includes some suppression of small counts and postprocessing for coterminous geography consistency but does not include Decennial Tabulation System suppressions. As a result, the SafeTab-H algorithm receives total population counts for more population groups than those released publicly in the Detailed DHC-A. Discrepancies between the counts input to SafeTab-H and the counts available in the Detailed DHC-A are strictly due to postprocessing of SafeTab-P's outputs. Importantly, these differences do not degrade the stated privacy guarantees of either SafeTab-P or SafeTab-H. Also, it should be noted that population groups that do appear in SafeTab-H's total population inputs and in the Detailed DHC-A have identical counts in both sources.

Because the DHC releases total housing unit counts without noise infusion, SafeTab-H does not produce statistics for geographic areas where no occupied or vacant housing units exist. This is ensured by removing such geographic areas from the input data before the SafeTab-H algorithm begins its processing.

6.5 Other Implementation Details

We note a few other implementation differences that are primarily driven by the specification requirements of the system and data wrangling aspects of the code. These include the function mapping detailed race and ethnicity codes to characteristic iterations, rules for what statistics are tabulated for different population groups, and the handling of Puerto Rico.

6.5.1 Mapping Detailed Race and Ethnicity Codes to Characteristic Iterations

The pseudocode in Section 3 abstracts the process of mapping a household's input record into their corresponding population group as a function g_i . In practice, this process requires joining against several specification input files and some subtle logic (to determine whether a household qualifies for an alone race characteristic iteration in addition to an alone or in any combination race characteristic iteration). However, the result is functionally equivalent to the g_i abstraction - each record is mapped to a number of geographic entities and characteristic iterations. The stability factor of the implemented transformations, the equivalent of $\Delta(g_i)$, is automatically tracked by Tumult Analytics rather than being computed by hand.

The pseudocode also ignores the logic associated with pre-processing a specified universe of geographic entities and iteration codes into population group levels \mathcal{P}_i . That is, the master list of all population groups divided into population group levels is constructed through a combination of specification files rather than being handed directly to the system.

6.5.2 Puerto Rico

The SafeTab-H algorithm presented in Algorithm 1 describes an input dataframe consisting of records of every household in the United States. However, the same algorithm is applied to a dataframe consisting of records from Puerto Rico. In implementation, SafeTab-H tabulates data for the United States and Puerto Rico in two separate passes.

7 Parameters and Tuning

Between the pseudocode representation of SafeTab-H and the selected implementation details presented earlier, we have alluded to a number of parameters that must be set before executing a run of the SafeTab-H program. Parameters are adjustable factors that must be fixed to fully define the nature of the program (e.g., the noise scale employed in `VECTORDISCRETEGAUSSIAN`, the privacy-loss budgets for population group levels, and total population thresholds to qualify for different household type and tenure table variants). With regard to SafeTab-H specifically (but any differentially private algorithm generally), parameter selection is a matter of policy. The Census Bureau’s Data Stewardship Executive Policy (DSEP) committee, in consultation with subject-matter experts as well as internal and external privacy experts, approved all available parameters for the Detailed DHC-B. Parameter selection necessitates trade-offs, as many of these parameters are dependent on each other. To illustrate, we consider a fundamental relationship between the privacy-loss parameters and their corresponding margins of error with discrete Gaussian noise distributions.

7.1 Error Bounds

SafeTab-H was designed to have predictable, tunable error so that accuracy targets of Census Bureau can be achieved with a known probability. Here, we examine the utility of Algorithm 1 with discrete Gaussian noise.

Definition 9. The 95% MOE is half the width of the 95% confidence interval.

Since 95% is the only confidence interval considered in this paper, we often write MOE without the 95% qualifier.

We begin by stating a portion of Proposition 25 from [9].

Proposition 1 (Proposition 25 in [9]). *For all $m \in \mathbb{Z}$ with $m \geq 1$, and for all $\sigma \in \mathbb{R}$ with $\sigma > 0$, $\Pr[X \geq m]_{X \leftarrow \mathcal{N}_{\mathbb{Z}}(\sigma^2)} \leq \Pr[X \geq m - 1]_{X \leftarrow \mathcal{N}(\sigma^2)}$.*

This says that discrete Gaussian tails are tighter than the corresponding continuous Gaussian tails, which follows because the discrete Gaussian is sub-Gaussian and has tighter variance.

The following corollary extends the tail bounds to the real numbers.

Corollary 1. *For all $x, \sigma \in \mathbb{R}$ with $x \geq 1$ and $\sigma > 0$, $\Pr[X > x]_{X \leftarrow \mathcal{N}_{\mathbb{Z}}(\sigma^2)} \leq \Pr[X > \lfloor x \rfloor]_{X \leftarrow \mathcal{N}(\sigma^2)}$.*

Figure 2 of [9] provides an intuitive visualization of these tail bounds. Using the continuous Gaussian to upper bound MOE in the discrete Gaussian, it follows that the discrete Gaussian has $X \in [-\lfloor 1.96\sigma \rfloor, \lfloor 1.96\sigma \rfloor]$ with probability of at least 95%. That is, $MOE \leq \lfloor 1.96\sigma \rfloor$.

Recall that $\sigma^2 = \frac{1}{2\rho}$ in Algorithm 3. Combining these two equations and solving for ρ , yields the following result.

Corollary 2. *The base discrete Gaussian mechanism run with $\rho^t = \frac{1.92}{\lfloor MOE \rfloor^2}$ has a 95% margin of error of at most MOE.*

For a population group in level i , the MOE in any single directly computed estimate in Algorithm 2 is $\left\lfloor 1.96 \sqrt{\frac{s}{2\rho_i^t}} \right\rfloor$ where $t \in \{\text{HT}, \text{T}\}$ and s is the stability of the function that maps records to their corresponding population groups in population group level i .

7.2 Parameter Identification, Trade-offs, and Outcomes

Parameters tend to impact some combination of these three aspects: data confidentiality, data accuracy, and data availability. Data availability refers to the volume of tabular statistics released. Data accuracy is primarily measured by the MOEs of the tabular statistics. Data confidentiality refers to privacy loss, and it is measured by the ρ -zCDP parameters of the algorithm. For example, excluding population group level i would reduce data availability but improve privacy since the privacy-loss budgets ρ_i^{HT} and ρ_i^{T} are no longer necessary.

To aid in the understanding of these trade-offs, we created the SafeTab-H Analysis Tool to provide hands-on experience exploring the parameter space. First, we provide a brief summary of this tool. Then, we highlight specific parameters and the critical decisions made by the DSEP committee based on recommendations from subject-matter expert or DAS scientists. Several of the subject-matter expert recommendations were influenced by interaction with the SafeTab-H Analysis tool.

7.3 Parameter Tuning Using the SafeTab-H Analysis Tool

The SafeTab-H Analysis Tool is an easy-to-use interactive decision support tool. The tool, implemented in the Microsoft Excel program, allowed users to interactively specify:

- The set of geography levels and characteristic iteration levels that constitute the universe of population groups for which statistics are tabulated.
- The maximum number of races a person is associated with to an integer in the range 1-8.
- Expected MOE targets with adjustable confidence levels.

Based on these parameters, the tool computed required privacy-loss parameters to achieve the desired target error. The computations were performed using analytical formulae for expected error of noise mechanisms employed in the SafeTab-H algorithm.

In the next section, we describe in the next section some of the key parameters considered and the decision process used to set these parameters.

7.3.1 Parameter Selection

Population group levels and population thresholds: As a reminder, a population group level is defined by a geographic summary level, such as Nation, State, and County, and an iteration level (i.e., Detailed or Regional). The SafeTab-H Analysis Tool helped subject-matter experts understand the privacy-loss budget required to produce acceptably accurate statistics at each geography level. The Census Bureau also gathered feedback from data users on these topics and ultimately settled on the levels referenced in Table 9. The population thresholds selected for Detailed DHC-B are similar to the population thresholds in Detailed DHC-A.

Race Multiplicity: The stability $\Delta(g_i)$ of the flatmap transformation g_i mapping households to population groups in level i is a significant factor in the noise scale required to satisfy zCDP. The data collection process restricts householders, and hence households, to a maximum of eight detailed race codes and one ethnicity code, which translates to a flatmap stability of nine for any given population group level. This is because an individual with eight unique detailed race codes can be associated with at most eight alone or in any combination characteristic iterations for a

Population Group Level	MOE Target	Unbounded Privacy Loss		Bounded Privacy Loss	
		Household Type	Tenure	Household Type	Tenure
(Nation, Detailed):	3	1.92	1.92	3.84	3.84
(State, Detailed):	3	1.92	1.92	3.84	3.84
(County, Detailed):	11	0.14	0.14	0.28	0.28
(Tract, Detailed):	11	0.14	0.14	0.28	0.28
(Place, Detailed):	11	0.14	0.14	0.28	0.28
(AIANNH, Detailed):	11	0.14	0.14	0.28	0.28
(Nation, Regional):	50	0.0069	0.0069	0.0138	0.0138
(State, Regional):	50	0.0069	0.0069	0.0138	0.0138
(County, Regional):	50	0.0069	0.0069	0.0138	0.0138
(Tract, Regional):	50	0.0069	0.0069	0.0138	0.0138
(Place, Regional):	50	0.0069	0.0069	0.0138	0.0138

Table 9: MOE targets for the statistics released at different population group levels, along with the corresponding privacy loss (unbounded and bounded ρ -zCDP for discrete Gaussian). The privacy loss is reported for both household type and tenure. Due to the total population of the United States being published without noise, the bounded privacy-loss budgets in this table were stressed by Census Bureau staff in internal conversations and for presentation to DSEP, for purposes of interpreting the privacy guarantee.

level plus the one ethnic characteristic iteration. Higher stability equates to higher variance noise, all else held equal, so an option to improve the noise variance would be to reduce the stability by setting a lower cap on the number of detailed race codes processed for each householder. For example, if householders were restricted to three race codes instead of eight, the stability would drop from nine to four, resulting in a 33% decrease in MOE when holding the privacy loss constant. However, the restriction would also introduce another form of bias into the statistics and potentially artificially reduce the set of population groups with true positive counts. The DSEP committee opted to leave the race multiplicity parameter at eight.

MOE and ρ : The SafeTab-H Analysis tool provided an interface for adjusting expected MOE targets to observe the impact on privacy loss, as derived in Section 7.1. The Census Bureau selected the MOEs and corresponding ρ s as displayed in Table 9 for the production run of SafeTab-H on the 2020 Census data.

8 Conclusion

In this paper, we presented the SafeTab-H algorithm, a differentially private algorithm for producing the Detailed DHC-B for the Census Bureau. We covered several key aspects of the algorithm. First, we provided a technical pseudocode description of the algorithm. Then, we covered the privacy properties of the algorithm. Next, we discussed salient differences between the pseudocode and the practical implementation of the algorithm with Tumult Analytics. We covered the key algorithm parameter choices made by Census Bureau policy and the ways in which the SafeTab-H Analysis Tool supported that process.

References

- [1] 2020 Census Operational Plan. <https://www2.census.gov/programs-surveys/decennial/2020/>
- [2] Rachel Marks and Merarys Rios-Vargas. Improvements to the 2020 Census Race and Hispanic Origin Question Designs, Data Processing, and Coding Procedures. <https://www.census.gov/newsroom/blogs/random-samplings/2021/08/improvements-to->
- [3] 2020 Census Detailed Demographic and Housing Characteristics File A (Detailed DHC-A) Technical Documentation. <https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation>
- [4] U.S. Code Title 13—Census. <https://www.law.cornell.edu/uscode/text/13>.
- [5] Seth Borenstein. Potential Privacy Lapse Found in Americans’ 2010 Census Data. <https://apnews.com/article/aba8e57c145047b5bab11b62baaa7f7a>, February 2019.
- [6] Memorandum 2019.25: 2010 Demonstration Data Products – Design Parameters and Global Privacy-Loss Budget. <https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-m> October 2019.
- [7] Frank McSherry. Privacy Integrated Queries: An Extensible Platform for Privacy-preserving Data Analysis. In Ugur Çetintemel, Stanley B. Zdonik, Donald Kossmann, and Nesime Tatbul, editors, *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2009, Providence, Rhode Island, USA, June 29 - July 2, 2009*, pages 19–30. ACM, 2009.
- [8] Mark Bun and Thomas Steinke. Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds. *CoRR*, abs/1605.02065, 2016.
- [9] Clément L. Canonne, Gautam Kamath, and Thomas Steinke. The Discrete Gaussian for Differential Privacy. *CoRR*, abs/2004.00010, 2020.
- [10] Skye Berghel, Philip Bohannon, Damien Desfontaines, Charles Estes, Sam Haney, Luke Hartman, Michael Hay, Ashwin Machanavajjhala, Tom Magerlein, Gerome Miklau, Amritha Pai, William Sexton, and Ruchit Shrestha. Tumult Analytics: A Robust, Easy-to-Use, Scalable, and Expressive Framework for Differential Privacy. <https://arxiv.org/abs/2212.04133>, 2022.