THE UNIVERSITY OF NEW MEXICO

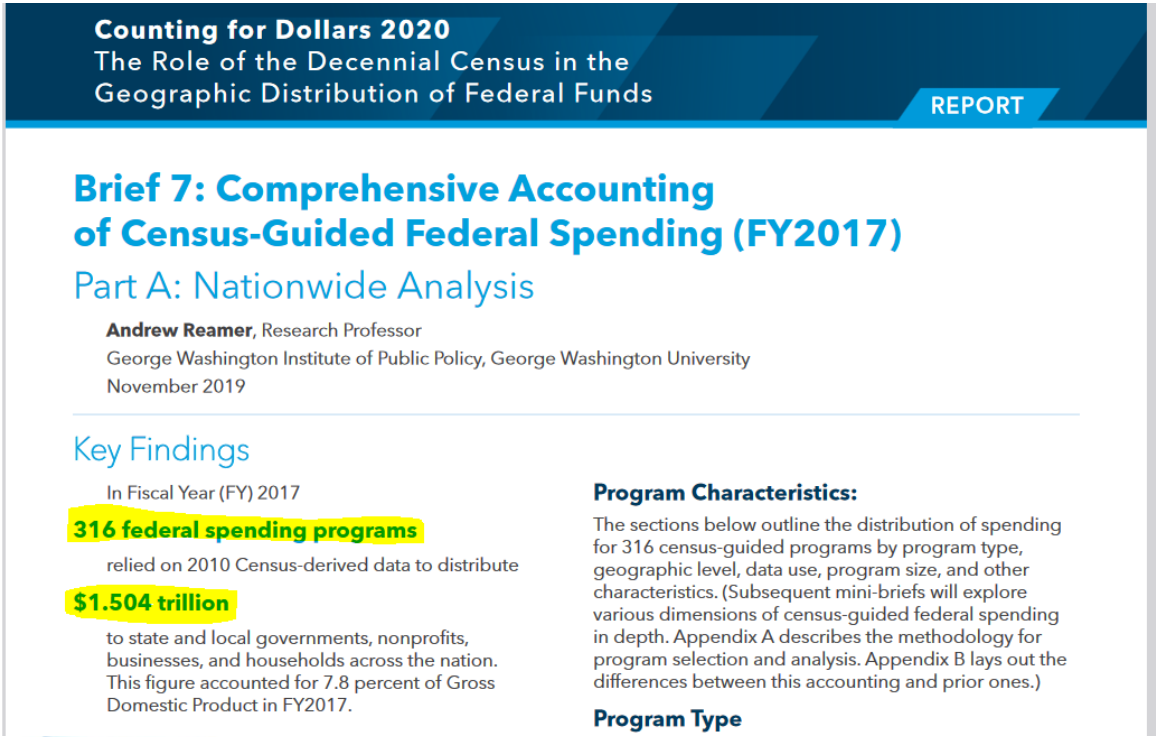
Notes on the Decennial Census

Robert Rhatigan
Director
Geospatial and Population Studies
University of New Mexico
rhatigan@unm.edu
505.277.4034

25th Annual New Mexico Data User's Conference
November 15, 2023

Why the Census Still Matter$

- Data
- Power
- Money


Counting for Dollars 2020
The Role of the Decennial Census in the Geographic Distribution of Federal Funds
REPORT

Brief 7: Comprehensive Accounting of Census-Guided Federal Spending (FY2017)

Part A: Nationwide Analysis

Andrew Reamer, Research Professor
George Washington Institute of Public Policy, George Washington University
November 2019

Key Findings

In Fiscal Year (FY) 2017

316 federal spending programs

relied on 2010 Census-derived data to distribute

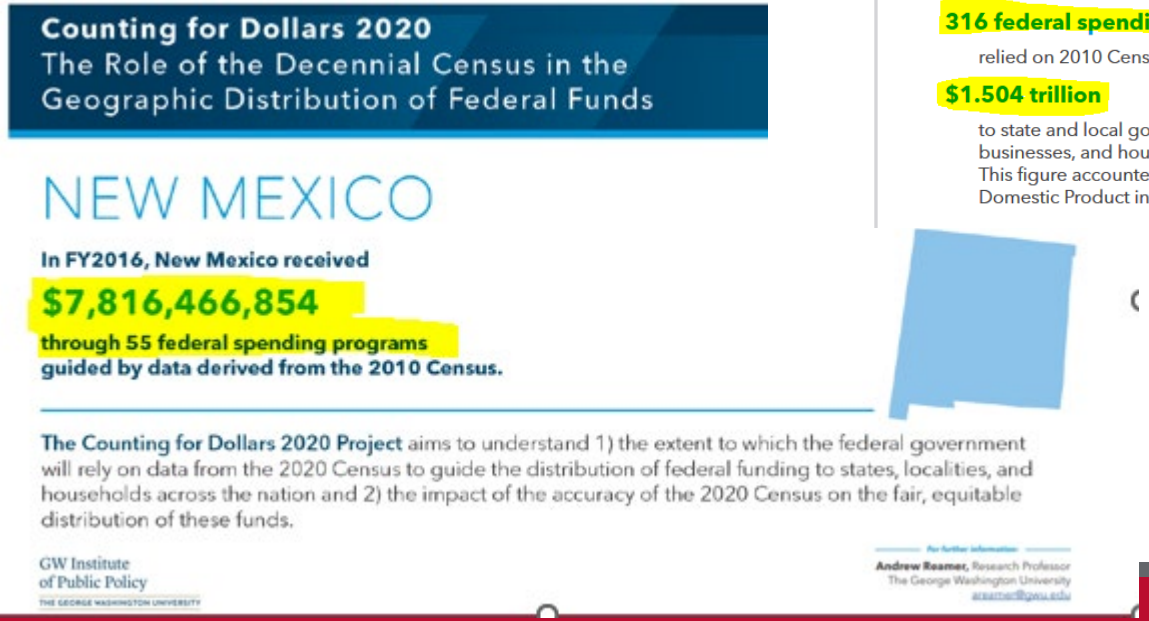
$1.504 trillion

to state and local governments, nonprofits, businesses, and households across the nation. This figure accounted for 7.8 percent of Gross Domestic Product in FY2017.

Program Characteristics:

The sections below outline the distribution of spending for 316 census-guided programs by program type, geographic level, data use, program size, and other characteristics. (Subsequent mini-briefs will explore various dimensions of census-guided federal spending in depth. Appendix A describes the methodology for program selection and analysis. Appendix B lays out the differences between this accounting and prior ones.)

Program Type


Counting for Dollars 2020
The Role of the Decennial Census in the Geographic Distribution of Federal Funds

NEW MEXICO

In FY2016, New Mexico received

$7,816,466,854

through 55 federal spending programs guided by data derived from the 2010 Census.

The Counting for Dollars 2020 Project aims to understand 1) the extent to which the federal government will rely on data from the 2020 Census to guide the distribution of federal funding to states, localities, and households across the nation and 2) the impact of the accuracy of the 2020 Census on the fair, equitable distribution of these funds.

GW Institute of Public Policy

Andrew Reamer, Research Professor
The George Washington University

Source: Counting for Dollars 2020, George Washington University Institute of Public Policy, The Role of the Decennial Census in the Geographic Distribution of Federal Funds; https://gwipp.gwu.edu/counting-dollars-2020-role-decennial-census-geographic-distribution-federal-funds

Outline

- **The Ongoing 2020 Census**
  - Emphasis on Disclosure Avoidance/Privacy Protections
- **Implications for New Mexico's Data Users**
  - Delays
  - The problem with small-area data
  - The population base
- **Plans for 2030 Census**
  - Administrative Records
  - Master Address File

Every Census is flawed, some more than others

Sources of Error:

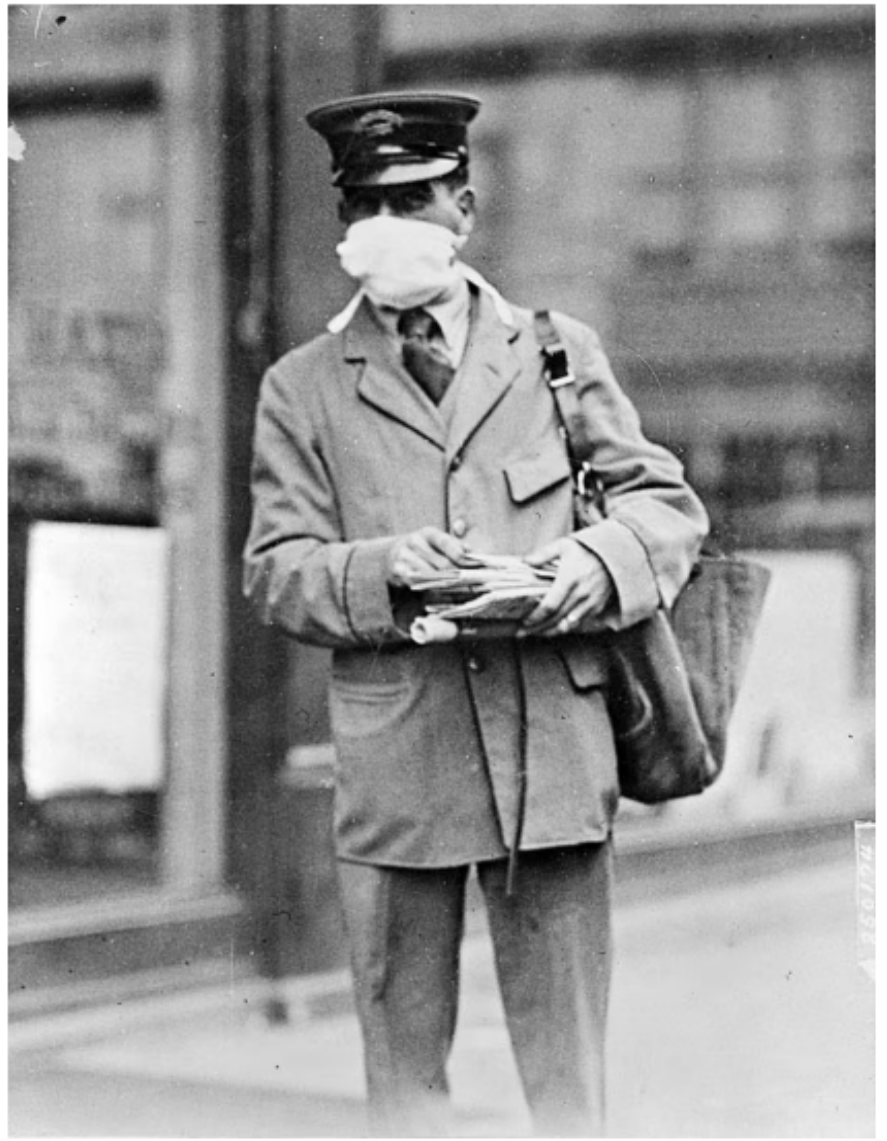
- **Omissions**
  - **including non-response**
- **Erroneous enumerations**
  - **including duplicates**
- **Natural disasters**
  - **Hurricanes**
  - **Wildfires**
  - **Pandemics**
- **Privacy protection measures**

U.S. Census Bureau History: The 1918 Influenza Pandemic

Doctors documented the "first" case of the H1N1 influenza pandemic on March 4, 1918, in a U.S. Army cook stationed at Camp Funston, Fort Riley, KS, named Albert Gitchell. Nicknamed the "Spanish Flu," cases of "severe influenza" were observed in Kansas as early as January 1918. Many scientists and researchers suspect that these Kansas infections were the first of an influenza pandemic that afflicted 500 million and killed as many as 100 million between 1918 and 1920 worldwide. The 1918 influenza pandemic was one of the greatest pandemics in human history with scenes of mask-clad adults and children, social distancing, and overwhelmed hospitals reminiscent of the more recent COVID-19 pandemic.

As World War I-weary Americans welcomed the arrival of the New Year in 1918, a "flu-like" virus was rapidly infecting, spreading, and killing previously healthy people in rural Haskell County, KS. Dr. Loring Miner, a Haskell County physician and coroner who visited some of these desperately ill patients noted that their health deteriorated more rapidly than he had seen during previous influenza outbreaks. From his analysis of samples collected in the first months of 1918, he reported to the U.S. Public Health Service that a novel and severe strain of influenza appeared to be rapidly spreading in the communities he served.

The Public Health Service published Loring's findings in an April 5, 1918, Health Alert—"On March 30, 1918, the



Isolated cases of the "Spanish Flu" occurred in early January 1918, but the first major wave of the virus began in March 1918. In scenes reminiscent of the recent COVID-19 pandemic, people going out into public wore a masks to protect against the deadly virus.

https://www.census.gov/history/www/homepage_archive/2023/march_2023.html

The Census Bureau could not have prepared to count 330,000,000 million people during the first wave of Covid-19

https://www.nj.gov/labor/labormarketinformation/demographics/us-census-data/2020census.shtml

Wins:

- A robust internet response option
- Committed and flexible career staff
- Brave field enumerators

+ Challenges:

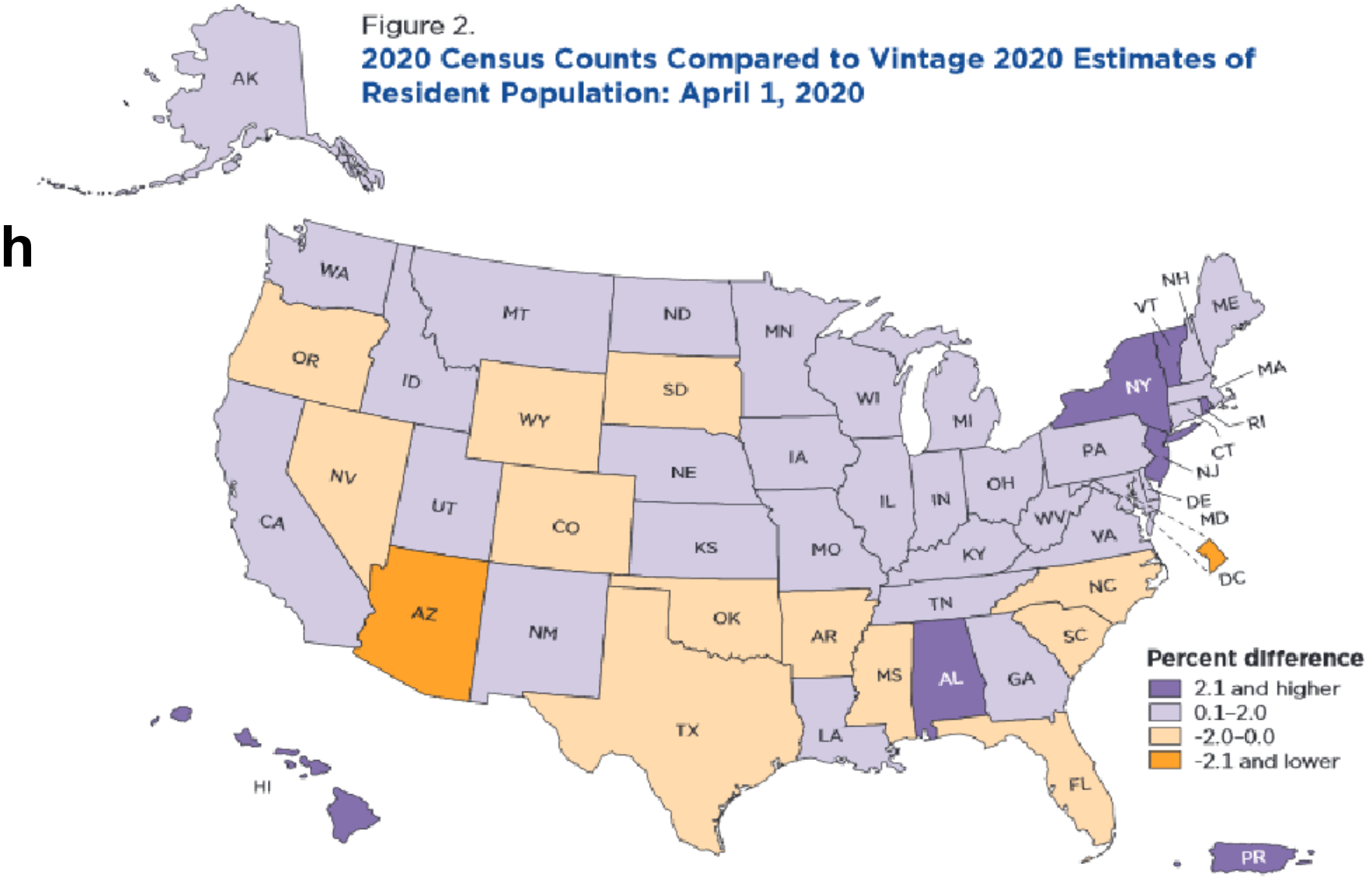
- Non-Response Follow Up
- Many people unwilling to answer their door
- Hiring and onboarding was difficult
- Personal Protective Equipment and training

= Field enumeration was extended by 2.5 months

https://www.chandleraz.gov/blog/covid-19s-impact-2020-census

Comparing the 2020 Census with 2020 Estimates

Figure 2.
2020 Census Counts Compared to Vintage 2020 Estimates of Resident Population: April 1, 2020

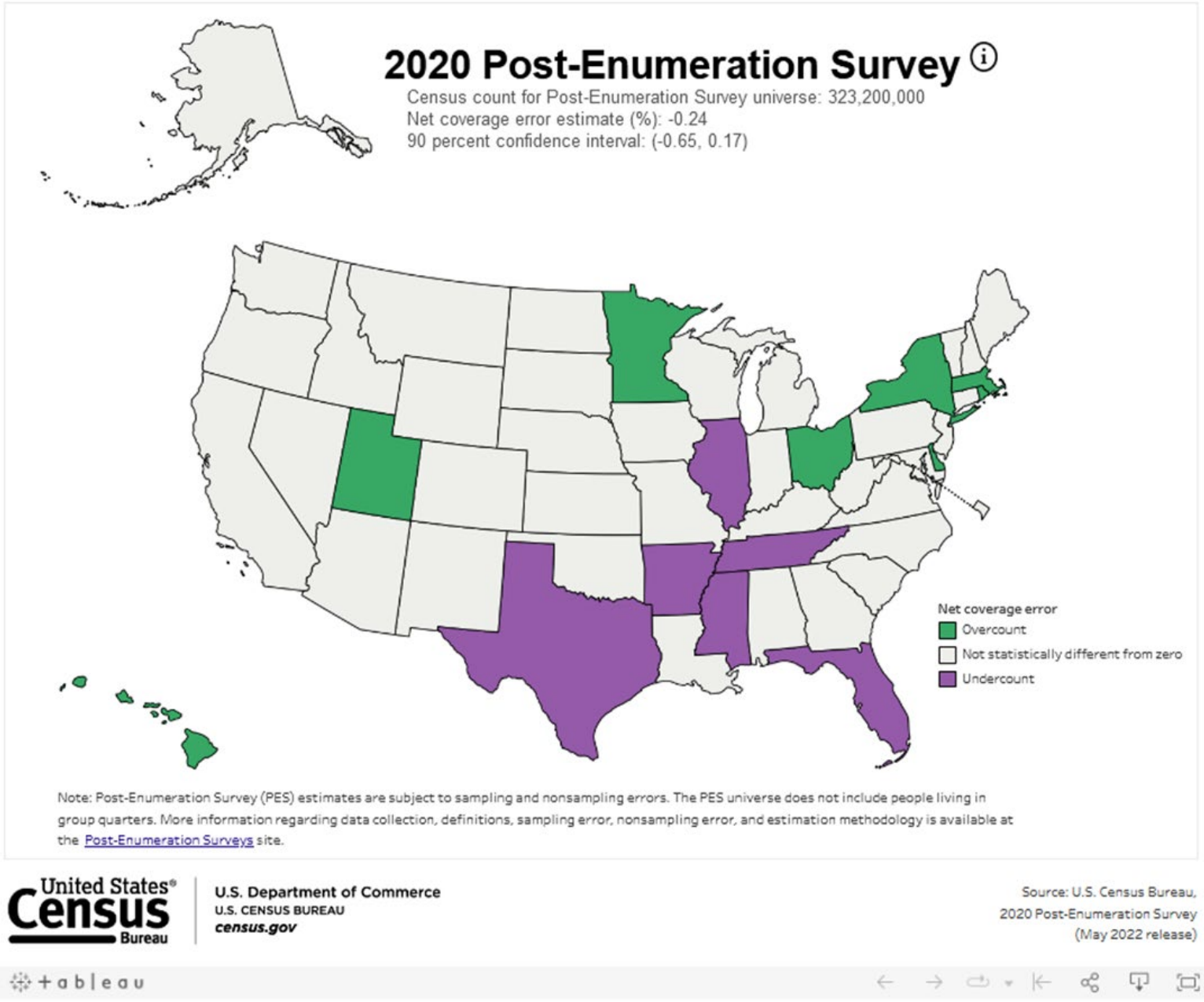


Source: U.S. Census Bureau, 2020 Census and Vintage 2020 Population Estimates.

2020 Post-Enumeration Survey Net Error by State

"In regards to an accurate count of the number of residents in a given state, NM had the most accurate count in the nation. However, nationwide, the demographic makeup of US residents undercounted Hispanics, American Indians, and Blacks. While we don't know what if any demographic-based undercount happened in NM, we can presume that there was some undercounting of certain demographics and some overcounting of others."

New Mexico Department of Finance and Administration Press Release June 8, 2022



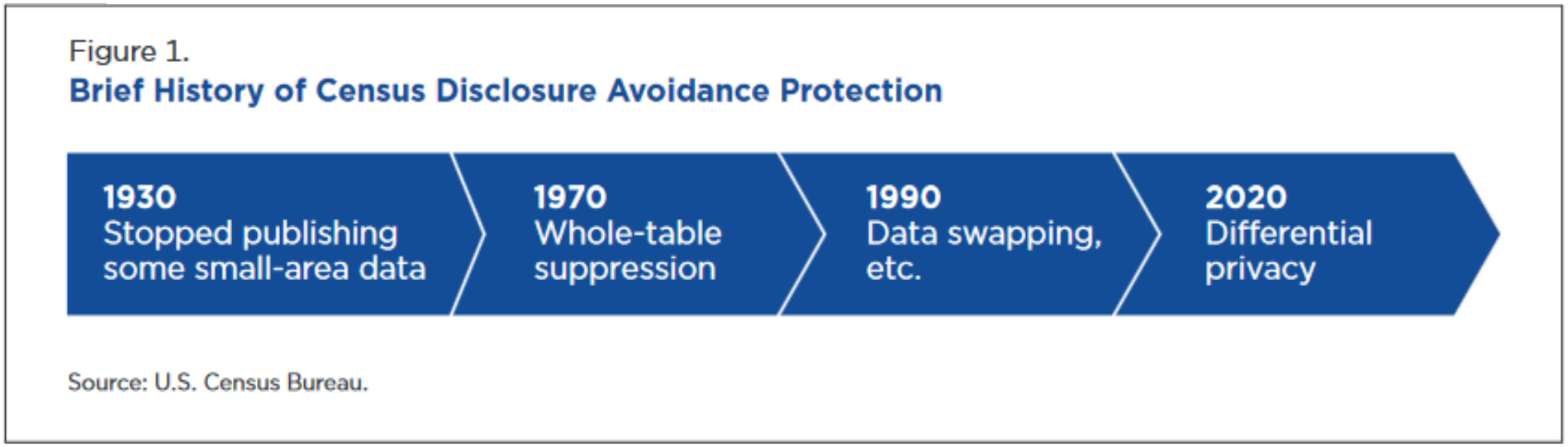
2020 Post-Enumeration Survey Net Error by State

	A	B	C	D	E
1	State	Census Count	Net Coverage Error (%)	Standard Error (%)	Absolute Net Error
2	New Mexico	2,075,000	0.10	2.42	0.10
3	New Jersey	9,109,000	-0.12	0.93	0.12
4	Connecticut	3,498,000	-0.13	1.71	0.13
5	Michigan	9,856,000	0.14	0.68	0.14
6	Arizona	6,991,000	-0.16	1.59	0.16
7	Indiana	6,607,000	0.21	0.94	0.21
8	Washington	7,545,000	0.29	1.08	0.29
9	Georgia	10,460,000	0.31	1.11	0.31
10	North Dakota	753,000	-0.35	1.86	0.35
11	Virginia	8,395,000	-0.35	1.11	0.35
12	South Dakota	855,000	-0.42	1.87	0.42
13	California	38,620,000	0.47	0.65	0.47
14	Pennsylvania	12,600,000	0.48	0.67	0.48
15	New Hampshire	1,335,000	0.53	1.58	0.53
16	Nebraska	1,912,000	-0.73	1.35	0.73
17	Kansas	2,851,000	-0.81	1.42	0.81
18	Wisconsin	5,742,000	0.85	0.86	0.85
19	Oregon	4,140,000	1.00	1.22	1.00
20	Alabama	4,896,000	-1.05	1.22	1.05
49	Arkansas	2,929,000	-5.04	2.21	5.04
50	Rhode Island	1,052,000	5.05	1.43	5.05
51	Delaware	967,000	5.45	2.82	5.45
52	Hawaii	1,415,000	6.79	1.68	6.79

Source Table:
https://data.census.gov/table/DECENNIALPES2020.G_STATES?g=010XX00US$0400000&y=2020&d=DEC+Decennial+Post-Enumeration+Survey

2020 Census Disclosure Avoidance

Title 13 of US Code requires the Census Bureau to protect the privacy of individual respondents



For the 2020 Census, the Census Bureau experimented with the modernization of Disclosure Avoidance

- Planning begin very late in the decennial cycle
- Key reason for delays in releasing data

From the National Academies of Sciences, Engineering, and Medicine, 2023. Assessing the 2020 Census: Final Report

“The Census Bureau decided to completely replace its methods for protecting the confidentiality of census data…with an entirely new approach…that had not been tested, prototyped, or deployed in the population census context.

In short, the new DAS was not ready for use in 2020 Census production and substantially degraded the value of 2020 Census data products in terms of both quality and timeliness.”

Source: National Academies of Sciences, Engineering, and Medicine. 2023. Assessing the 2020 Census: Final Report. Washington, DC: The National Academies Press. https://doi.org/10.17226/27150

National Academies of Sciences, Engineering, and Medicine, 2023. Assessing the 2020 Census: Final Report

“The Census Bureau decided to completely replace its methods for protecting the confidentiality of census data (termed Disclosure Avoidance System [DAS]) **with an entirely new approach—one that had not been tested, prototyped, or deployed in the population census context.** While confidentiality protection is a critically important responsibility of a statistical agency, this decision was made without appropriate consideration and balance regarding the utility of resulting census data products to fulfill the many important functions of census data. **In short, the new DAS was not ready for use in 2020 Census production and substantially degraded the value of 2020 Census data products in terms of both quality and timeliness.** Moreover, lingering questions about both the simulated database reconstruction attack that motivated the new DAS and the degree of confidentiality protection that was ultimately realized through final parameter settings have arguably harmed 2020 Census data products, some of which are not set to be released until late 2024, and the reputation of the Census Bureau.”

Source: National Academies of Sciences, Engineering, and Medicine. 2023. Assessing the 2020 Census: Final Report. Washington, DC: The National Academies Press. https://doi.org/10.17226/27150

2020 Census Small-Area Data

- Block-level census data has always been
  - error-prone
  - highly uncertain
  - subjected to privacy protections

- Block-level data were historically
  - consistent
  - aligned with reality

- 2020 block-level data can be illogical and not fit-for-use

Small-Area Data Problem Example #1

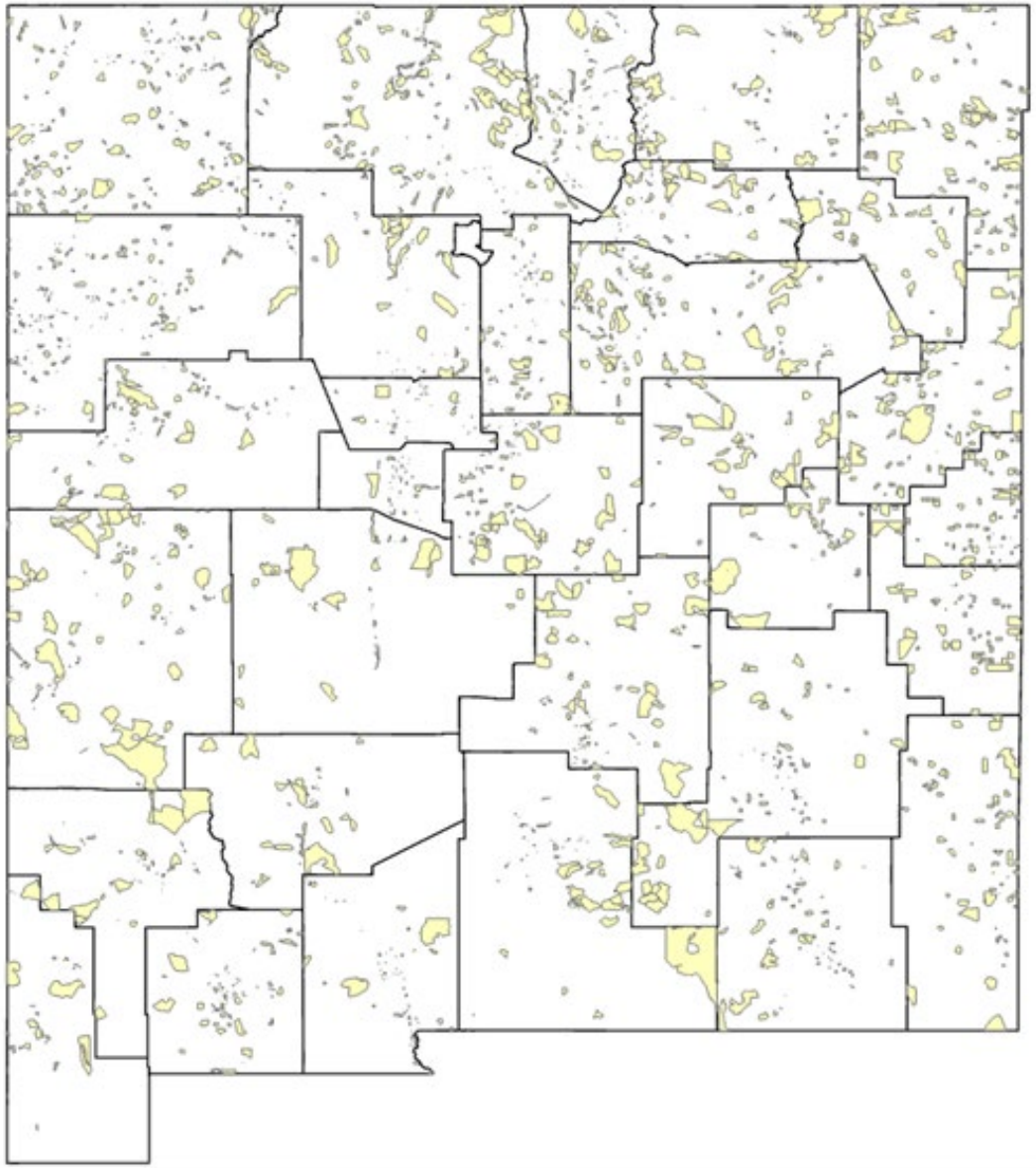
New Mexico census blocks with:

a population of one or more people *and*

no occupied housing units or group quarters

Representing:

- 5,647 blocks
- 48,682 people



Source: 2020 P.L. 94-171 Redistricting Data

Small-Area Data Problem Example #2

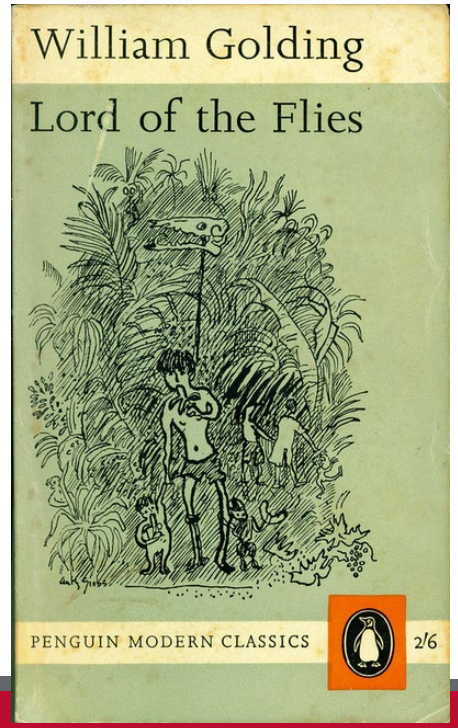
New Mexico Census Blocks with:

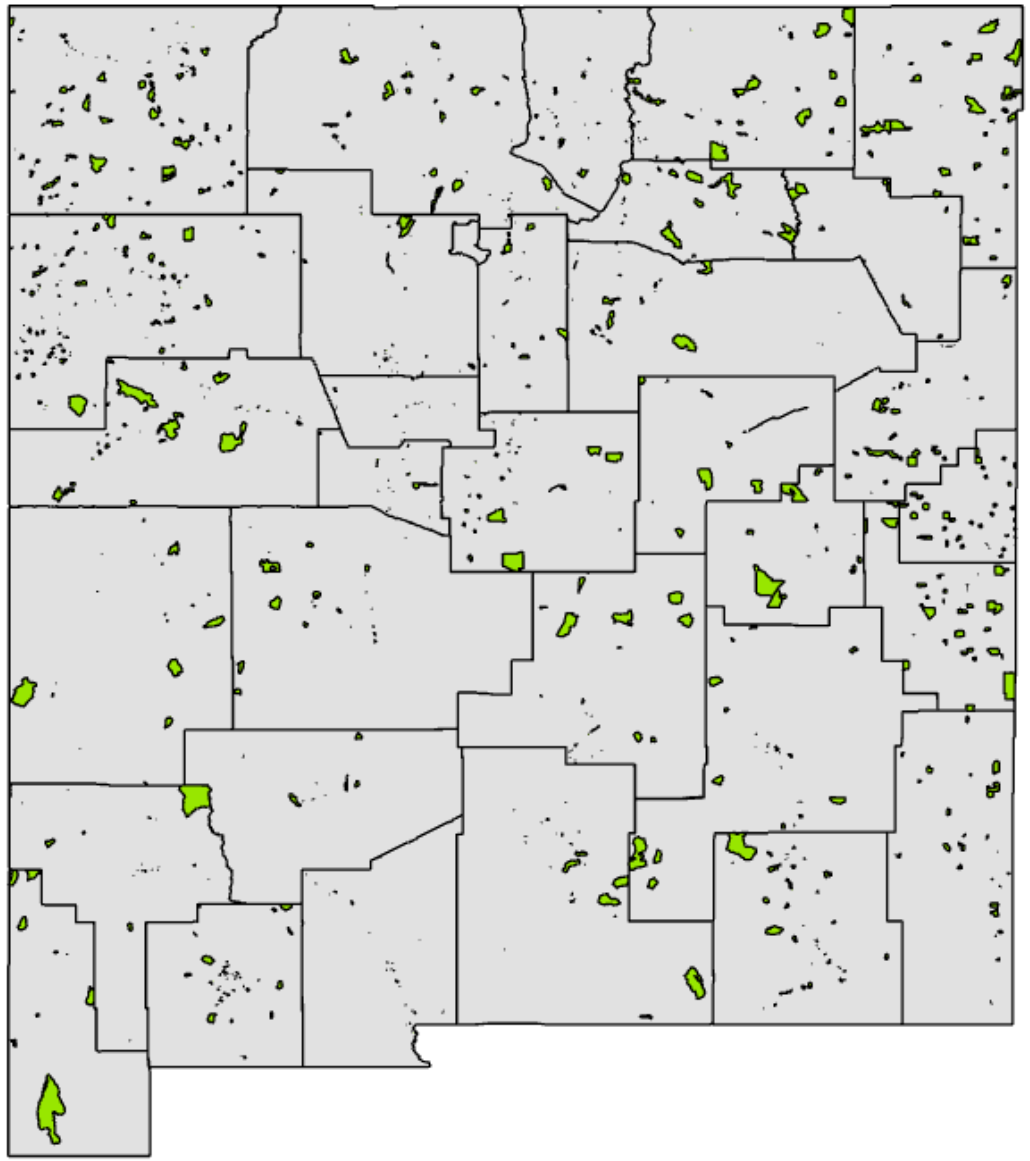
a population of one or more children ages 0-17 *and* no population of adults 18+

Representing:

-1,155 Blocks

-2,924 Children





Source: 2020 P.L 94-171 Redistricting Data

Aggregating Block-Level Data

To reduce error, the Census Bureau advises aggregating blocks to populations of 500 or more

Table 4.

Differences Between the TopDown and SafeTab-P Differential Privacy Algorithms

Algorithm	Privacy	Geography	Accuracy When Aggregating Data	Consistency Across Products	Accuracy Targets	Design
TopDown (Redistricting Data, DHC, Demographic Profile)	Algorithm produces privacy-protected microdata.	All geographies can be aggregated as expected.	When aggregating data, the statistical noise generally cancels out, and the statistics become more accurate.	Consistent across data products.	Overall, accuracy can be targeted but the exact levels of accuracy cannot be known in advance.	Does not use adaptive design.
SafeTab-P (Detailed DHC-A)	Algorithm directly produces privacy-protected tabulations.	There is no requirement that geographies can be aggregated as expected.	When aggregating data, it generally becomes more variable the more you aggregate.	Not consistent with other 2020 Census data products.	All margins of error are determined in advance and met 95 percent of the time.	Uses adaptive design to determine the amount of data provided.

Note: DHC is Demographic and Housing Characteristics.
Source: U.S. Census Bureau.

Aggregating Block-Level Data

BUT, only aggregate block with certain datasets, and not others.

Table 4.

Differences Between the TopDown and SafeTab-P Differential Privacy Algorithms

Algorithm	Privacy	Geography	Accuracy When Aggregating Data	Consistency Across Products	Accuracy Targets	Design
TopDown (Redistricting Data, DHC, Demographic Profile)	Algorithm produces privacy-protected microdata.	All geographies can be aggregated as expected.	When aggregating data, the statistical noise generally cancels out, and the statistics become more accurate.	Consistent across data products.	Overall, accuracy can be targeted but the exact levels of accuracy cannot be known in advance.	Does not use adaptive design.
SafeTab-P (Detailed DHC-A)	Algorithm directly produces privacy-protected tabulations.	There is no requirement that geographies can be aggregated as expected.	When aggregating data, it generally becomes more variable the more you aggregate.	Not consistent with other 2020 Census data products.	All margins of error are determined in advance and met 95 percent of the time.	Uses adaptive design to determine the amount of data provided.

Note: DHC is Demographic and Housing Characteristics.
Source: U.S. Census Bureau.

Implication: 2020 Census Data Release Delays

- ✓ Apportionment Data released April, 2021
- ✓ Redistricting data released August, 2021
- ✓ DHC – released May, 2023
- ✓ DHC-A released August, 2023
- ? DHC-B **release TBD**
- ? Supplemental-DHC, **release TBD**

https://www.census.gov/newsroom/press-releases/2022/2020-census-data-products-schedule-2023.html

Implication: Small-Area Data

Small-Area and Small-Cohort data are highly uncertainty

-This has always been the case

-It's a bigger problem now

***Custom geographies* are problematic in small communities e.g.**

-Legislative Districts

-School Districts

-Business Planning/Market Analyses

-Transportation Planning

-Soil and Water Conservation Districts

Implication: The Blended Base

- Each decade, a new population base is derived from the census
- This population base is the foundation for a decade of Population Estimates
- The base is historically established in year 1 of each decade
- In 2023, we have an evolving “Blended Base”
  - This is a creative solution developed with input from external stakeholders
- 2020 Census data is “blended” with other data sources specifically
  - 2010-based population estimates
  - Characteristics from Demographic Analysis

For Blended Base see: https://www.census.gov/library/stories/2023/06/blended-base-methodology.html

Image:https://commons.wikimedia.org/wiki/File:Kitchen_blender.svg

Implications for Other Data

Uncertain, questionable, and erroneous census data effects many other demographic dataset e.g.

- **Population Estimates** – used for distribution of $1.5 trillion in annual federal funds
- **Population Projections** – critical for community planning
- **American Community Survey** – largest sample survey
- **Current Population Survey** – unemployment rates
- **Burea of Labor Statistics/Bureau of Economic Analysis**
- **Practically any person-level sample survey** is impacted because the population control totals are derived from the decennial census

*Note that privacy purists want to layer privacy protects i.e. apply privacy protection to every data product even those derived from data that already have privacy protections in place.

Census Bureau Plans for Disclosure Avoidance Systems

The Top-Down Algorithm (TDA), the original implementation of Differential Privacy, is not being used for all 2020 Census products

- The SafeTab-P Algorithm was used for DHC-A
- DHC-B and Supplemental DHC will not use TDA.
- The 2022 Economic Census will continue to use cell suppression, not Differential Privacy

To Clarify:

My criticisms of Disclosure Avoidance in the 2020 Census relate to a lack of planning and insufficient time for testing and iteration.

An elegant theory, Differential Privacy, was rushed into production using America's flagship data; the decennial census.

This was not a step forward towards modernization of privacy protection

2030 Census Planning

Lest we forget that New Mexico is a Hard-to-Count state with many Historically Undercounted Populations

Hard-To-Count Populations

- Young children
- Racial and ethnic minorities
- People who speak languages other than English
- Low-income populations
- Undocumented immigrants
- LGBTQ+
- People with disabilities
- People who do not live in traditional housing

Historically Undercounted Populations

- Black or African American population
- American Indian and Alaska Native population living on reservations
- People who indicate that they are some other race than the categories offered
- Hispanic or Latino population
- Young children, ages 0-4
- Renters
- Males, ages 18-29 and ages 30-49

See: https://www.census.gov/newsroom/blogs/random-samplings/2023/10/understanding-undercounted-populations.html

Census Bureau 2030 Research Agenda

Includes:
✓ Expanding the use of Administrative Records

Does not Include:
X Disclosure Avoidance
X **Improving the Master Address File**

New Mexico's address list is among the most problematic in the county
Tribal Lands
Rural areas without home mail delivery e.g. Mora County, Catron County

A poor address list effects the Census as well as
- Emergency Response
- Epidemiology
- Taxation
- **Broadband Access and Expansion**

Questions, Comments, Rebukes?

Robert Rhatigan
Director
Geospatial and Population Studies
University of New Mexico
rhatigan@unm.edu
505.277.4034

25th Annual New Mexico Data User's Conference
November 15, 2023