IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

ELECTRONIC PRIVACY INFORMATION CENTER, Plaintiff, v. UNITED STATES DEPARTMENT OF COMMERCE, et al., Defendants.	Civ. Action No. 18-2711 (DLF)

PLAINTIFF'S MOTION FOR A PRELIMINARY INJUNCTION

Pursuant to Federal Rule of Civil Procedure 65 and Local Rule 65.1(c), Plaintiff Electronic Privacy Information Center ("EPIC") respectfully moves the Court to issue a preliminary injunction preventing Defendants United States Department of Commerce and Bureau of the Census from (1) implementing the Defendants' March 26, 2018 decision to add a citizenship question to the 2020 Census, and (2) otherwise initiating any collection of citizenship status information that would be obtained through the 2020 Census. This injunction is necessary to preserve the status quo pending final adjudication of EPIC's claims under the Administrative Procedure Act and the E-Government Act, Pub. L. No. 107-347, 116 Stat. 2899 (Dec. 17, 2002) (codified at 44 U.S.C. § 3501 note).

The full grounds for this motion are set out in the accompanying Memorandum in Support of Plaintiff's Motion for a Preliminary Injunction. Pursuant to Local Rule 65.1(d), and for the reasons set forth the Memorandum, EPIC requests that the Court schedule a hearing on this motion at the Court's earliest convenience.

Respectfully Submitted,

MARC ROTENBERG, D.C. Bar #422825
EPIC President and Executive Director

ALAN BUTLER, D.C. Bar #1012128
EPIC Senior Counsel

/s/ John Davisson
JOHN DAVISSON, D.C. Bar #1531914
EPIC Counsel

ELECTRONIC PRIVACY
INFORMATION CENTER
1718 Connecticut Avenue, N.W.
Suite 200
Washington, D.C. 20009
(202) 483-1140 (telephone)
(202) 483-1248 (facsimile)

Attorneys for Plaintiff EPIC

Dated: January 18, 2018

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

ELECTRONIC PRIVACY INFORMATION CENTER,

Plaintiff,

v.

UNITED STATES DEPARTMENT OF COMMERCE, et al.,

Defendants.

Civ. Action No. 18-2711 (DLF)

MEMORANDUM IN SUPPORT OF PLAINTIFF'S MOTION FOR A PRELIMINARY INJUNCTION

TABLE OF CONTENTS

TABLE OF AUTHORITIES ii
SUMMARY 1
BACKGROUND 2
I. The E-Government Act of 2002 2
II. The Department of Commerce order initiating the collection of personal data concerning citizenship status 3
III. The privacy implications of collecting citizenship status information 6
IV. The lack of privacy impact assessments analyzing the citizenship question 9
V. EPIC's prior warnings to the Census Bureau of E-Government Act noncompliance 12
VI. EPIC's lawsuit under the E-Government Act 13
VII. The ruling of the court in *New York v. U.S. Department of Commerce* 14
ARGUMENT 15
I. EPIC is entitled to a preliminary injunction. 15
A. EPIC is likely to succeed on the merits of its claims. 16
1. The Census Bureau is required to publish privacy impact assessments that fully address the privacy risks of collecting citizenship status information 17
2. The Census Bureau was obligated to publish the required privacy impact assessments by March 26, 2018. 19
3. The Census Bureau has unlawfully failed to publish the required privacy impact assessments for more than nine months. 21
4. EPIC's suit satisfies the requirements for an action under the APA. 24
B. EPIC will suffer irreparable harm absent an injunction. 26
C. The equities and the public interest weigh in favor of an injunction. 33
II. EPIC has Article III standing in this matter. 37
CONCLUSION 40

TABLE OF AUTHORITIES

Cases

Baldrige v. Shapiro,
455 U.S. 345 (1982) 1

Benisek v. Lamone,
138 S. Ct. 1942 (2018) 16

Benkelman Tel. Co. v. FCC,
220 F.3d 601 (D.C. Cir. 2000) 20

Byrd v. EPA,
174 F.3d 239 (D.C. Cir. 1999) 27

Calvert Cliffs' Coordinating Comm., Inc. v. AEC,
449 F.2d 1109 (1971) 30

Chaplaincy of Full Gospel Churches v. England,
454 F.3d 290 (D.C. Cir. 2006) 29, 31

City of San Jose v. Ross,
No. 18-2279 (N.D. Cal. filed Apr. 17, 2018) 28

Clapper v. Amnesty Int'l USA,
568 U.S. 398 (2013) 38

Conference of State Bank Supervisors v. Office of Comptroller of Currency,
313 F. Supp. 3d 285 (D.D.C. 2018) 37, 38

Ctr. for Sustainable Econ. v. Jewell,
779 F.3d 588 (D.C. Cir. 2015) 38, 39

EPIC v. DOJ,
416 F. Supp. 2d 30 (D.D.C. 2006) 27

EPIC v. Presidential Advisory Comm'n on Election Integrity,
266 F. Supp. 3d 297 (D.D.C.), *aff'd on other grounds*, 878 F.3d 371 ..... 27, 28, 29, 33

EPIC v. Presidential Advisory Comm'n on Election Integrity,
878 F.3d 371 (D.C. Cir. 2017) 20, 25, 26, 30, 38

FEC v. Akins,
524 U.S. 11 (1998) 37, 38

Fed'n for Am. Immigration Reform v. Klutznick,
486 F. Supp. 564 (D.D.C. 1980) 8

Friedman v. FAA,
841 F.3d 537 (D.C. Cir. 2016).....26

Hirschfeld v. Stone,
193 F.R.D. 175 (S.D.N.Y. 2000).....32

Illinois Commerce Comm'n v. ICC,
848 F.2d 1246 (D.C. Cir. 1988).....30

In re Dep't of Commerce,
No. 18-557, 2018 WL 5458822 (U.S. Nov. 16, 2018).....15, 28

Int'l Union, United Auto., Aerospace & Agric. Implement Workers of Am. v. Brock,
477 U.S. 274 (1986).....39

Ipsen Biopharmaceuticals, Inc. v. Hargan,
334 F. Supp. 3d 274 (D.D.C. 2018).....25

Jones v. D.C. Redevelopment Land Agency,
499 F.2d 502 (D.C. Cir. 1974).....27, 30, 31, 34, 36

Judicial Watch, Inc. v. DHS,
895 F.3d 770 (D.C. Cir. 2018).....27

Kelso v. U.S. Dep't of State,
13 F. Supp. 2d 1 (D.D.C. 1998).....19

Kravitz v. Dep't of Commerce,
No. GJH-18-1041, 2018 WL 6830226 (D. Md. Dec. 28, 2018).....24, 28

La Union del Pueblo Entero v. Ross,
No. GJH-18-1570, 2018 WL 5885528 (D. Md. Nov. 9, 2018).....28

Lathan v. Volpe,
455 F.2d 1111 (9th Cir. 1971).....30, 31, 34, 36

Nat'l Fed'n of Fed. Emps. v. Greenberg,
789 F. Supp. 430 (D.D.C. 1992).....32

Navajo Nation v. Azar,
292 F. Supp. 3d 508 (D.D.C. 2018).....15, 26, 33

New York Immigration Coal. v. Dep't of Commerce,
No. 18-CV-5025 (JMF), 2019 WL 190285 (S.D.N.Y. Jan. 15, 2019).....28

New York v. U.S. Dep't of Commerce,
No. 18-CV-2921 (JMF), 2019 WL 190285 (S.D.N.Y. Jan. 15, 2019).....5, 8, 14, 24, 28

Nken v. Holder,
556 U.S. 418 (2009) 33

Norton v. Southern Utah Wilderness Alliance (*SUWA*),
542 U.S. 55 (2004) 25, 26

Payne Enters., Inc. v. United States,
837 F.2d 486 (D.C. Cir. 1988).....27

Pharm. Research & Mfrs. of Am. v. Thompson,
259 F. Supp. 2d 39 (D.D.C. 2003), *aff'd*, 362 F.3d 817 (D.C. Cir. 2004) 24

Pub. Employees for Envtl. Responsibility v. United States Fish & Wildlife Serv.,
189 F. Supp. 3d 1 (D.D.C. 2016) 31

Public Citizen v. DOJ,
491 U.S. 440 (1989) 38

Realty Income Tr. v. Eckerd,
564 F.2d 447 (D.C. Cir. 1977).....31, 35

Saco River Cellular, Inc. v. FCC,
133 F.3d 25 (D.C. Cir. 1998).....20

Sierra Club v. Thomas,
828 F.2d 783 (D.C. Cir. 1987).....26

Sorenson Commc'ns, LLC v. FCC,
897 F.3d 214 (D.C. Cir. 2018).....38

Spokeo, Inc. v. Robins,
136 S. Ct. 1540 (2016) 38

Steel Co. v. Citizens for a Better Env't,
523 U.S. 83 (1998) 37

Summers v. Earth Island Inst.,
555 U.S. 488 (2009) 37

U.S. Army Corps of Eng'rs v. Hawkes Co.,
136 S. Ct. 1807 (2016) 25

United States v. Hubbard,
650 F.2d 293 (D.C. Cir. 1980).....32

University of Tex. v. Camenisch,
451 U.S. 390 (1981) 16

W. Org. of Res. Councils v. Zinke,
892 F.3d 1234 (D.C. Cir. 2018).....25, 26

Washington Post v. DHS,
459 F. Supp. 2d 61 (D.D.C. 2006) 27

Winter v. NRDC,
555 U.S. 7 (2008) 15, 26, 29, 33

Youngstown Sheet & Tube Co. v. Sawyer,
343 U.S. 579 (1952) (Frankfurter, J., concurring).....35

Statutes

13 U.S.C. § 5 4

13 U.S.C. § 141(a).....4

13 U.S.C. § 221 .....4, 17, 21, 25

15 U.S.C. § 1501 4

28 U.S.C. § 2201(a).....14

42 U.S.C. § 4332 30

44 U.S.C. § 3502(3)(A) 20, 21

44 U.S.C. § 3902 19

Act of Jul. 2, 1909 (to provide for the expenses of the Thirteenth December Census, and for other purposes), ch. 2, § 25, 36 Stat. 1, 9 7

Act of June 18, 1929 (to provide for the fifteenth and subsequent decennial censuses and to provide for apportionment of Representatives in Congress), ch. 28, § 11, 46 Stat. 21, 25.....7

Administrative Procedure Act

5 U.S.C. § 702 24

5 U.S.C. § 704 .....24, 25

5 U.S.C. § 706 .....24, 25

5 U.S.C. § 706(1).....14, 26

5 U.S.C. § 706(2).....14, 25

E-Government Act, Pub. L. No. 107-347, 116 Stat. 2899 (Dec. 17, 2002) (codified at 44 U.S.C. § 3501 note)

§ 2(b) 2, 36

§ 201 20

§ 208(a) 25, 26, 29, 36

§ 208(b) 2, 3, 14, 16, 17, 19, 22, 23, 26, 39

§ 208(d) 3, 17

Regulations

5 C.F.R. § 1320.3(c) 20, 21

Other Authorities

148 Cong. Rec. 11,227 (2002) (statement of Sen. Lieberman) 2

Black's Law Dictionary 705 (Special Deluxe 5th ed. 1979) 19

Br. of EPIC as Amicus Curiae in Supp. Pls.' Position at Trial, *New York v. U.S. Dep't of Commerce*, No. 18-CV-2921 (JMF), 2019 WL 190285 (S.D.N.Y. Jan. 15, 2019) 13, 34

"Census" and "Citizenship", Google News (Jan. 13, 2019) 28

Census: Australian Bureau of Statistics Says Website Attacked by Overseas Hackers, ABC News (Aug. 9, 2016) 1

Decl. of Andrew Case in Supp. of Pls.' Opp'n to Defs.' Mot. Summ. J. at Ex. B, *San Jose v. Ross*, 18-2279 (N.D. Cal. Filed Nov. 16, 2018) 6

Defs.' Second Suppl. Resps. to Pls.' First Set of Interrogatories, *N.Y. Immigration Coal. v. U.S. Dep't of Commerce,* 18-5025 (Oct. 11, 2018) 5

Dep't of Commerce, Office of Privacy & Open Gov't, *U.S. Census Bureau Privacy Impact Assessments (PIAs) and Privacy Threshold Analysis (PTA)* (Oct. 1, 2018). 9

Email from Kris Kobach, Sec'y, Kan. Dep't of State, to Wilbur Ross, Sec'y, Dep't of Commerce (Jul. 21, 2017) 5

EPIC, *Department of Homeland Security Obtained Data on Arab Americans From Census Bureau* (2004) 8

EPIC, *EPIC FOIA Cases* (Sep. 17, 2018) 34, 39

Joshua B. Bolten, Dir., OMB, Executive Office of the President, M03-22, Memorandum for Heads of Executive Departments and Agencies, Attachment A (Sept. 26, 2003) ("OMB Guidance") 16, 17, 18, 19 22, 23, 38

Latanya Sweeney, *Simple Demographics Often Identify People Uniquely* (Carnegie Mellon Univ., Data Privacy Working Paper No. 3, 2000).....6, 18

Letter from Attorneys General of Twenty U.S. States to Wilbur Ross, Secretary of Commerce (Feb. 12, 2018) 5

Letter from Sen. Dianne Feinstein et al. to Wilbur Ross, Secretary of Commerce (Jan. 5, 2018).....5

Lynette Clemetson, *Census Policy On Providing Sensitive Data Is Revised*, N.Y. Times, (Aug. 31, 2004) 8

Lynette Clemetson, *Homeland Security Given Data on Arab-Americans*, N.Y. Times (Jul. 30, 2004).....8

Margo Anderson & William Seltzer, *Census Confidentiality Under the Second War Powers Act (1942- 1947)* (Mar. 29-31, 2007) (unpublished manuscript).....7

Margo Anderson & William Seltzer, *Challenges to the Confidentiality of U.S. Federal Statistics*, 1910-1965, 23 J Official Stat. 1, 6–7 (2007).....7

Petition for a Writ of Mandamus, *In re Dep't of Commerce*, No. 18-557, 2018 WL 5458822 (U.S. Nov. 16, 2018) 36

Presidential Proclamation No. 7,286, 65 Fed. Reg. 17,985 (Apr. 1, 2000).....2, 36

Proposed Information Collection; Comment Request; 2020 Census, 83 Fed. Reg. 26,643 (June 8, 2018).....12

Submission for OMB Review, 83 Fed. Reg. 67,213 (Dec. 28, 2018).....13, 21, 32, 35

Tara Bahrampour, *Trump Administration Officials Suggested Sharing Census Responses with Law Enforcement, Court Documents Show*, Wash. Post (Nov. 19, 2018).....6

U.S. Census Bureau, *2020 Census* (Oct. 19, 2018).....4

U.S. Census Bureau, *2020 Census Operational Plan: A New Design for the 21st Century* (Sept. 2017).....13, 28, 31, 37

U.S. Census Bureau, *Decennial Census of Population and Housing* (2014) 21

U.S. Census Bureau, *Subjects Planned for the 2020 Census and American Community Survey* (Mar. 2017) 21

U.S. Conference of Mayors, *Nation's Mayors to Secretary Ross: Don't Politicize Census. Remove the Citizenship Question* (Mar. 27, 2018) 5

U.S. Customs and Border Protection, *Policy for Requesting Information of a Sensitive Nature from the Census Bureau* (Aug. 9, 2004) 8

SUMMARY

Unique among federal agencies, the U.S. Census Bureau is authorized by law to compel, from every person in the United States, their personal data including age, sex, race, ethnicity, family relationships, and homeownership status. The extraordinary reach of the Bureau into the private lives of Americans brings with it extraordinary risks to privacy. The data collected by the Bureau is not only subject to uses by other federal agencies inconsistent with the collection purpose, but may also be the target of foreign cyberattacks.[1] Accordingly, "Congress has provided assurances that information furnished to the [Census Bureau] by individuals is to be treated as confidential." *Baldrige v. Shapiro*, 455 U.S. 345, 354 (1982). These legal assurances include the transparency and oversight provisions of the E-Government Act, which require the Census Bureau, a component of the Department of Commerce, to conduct and publish privacy impact assessments *before* initiating the collection of personal data.

But the Census Bureau has failed to live up to its obligations. The Bureau decided to add a citizenship question to the census yet refused to evaluate the privacy threats posed by the collection of citizenship status information, as required by the E-Government Act. As a result of the failure to complete mandatory privacy impact assessments, the Bureau has denied EPIC access to information vital to the important public debate over census changes, has undermined the agency's own decisionmaking process, and has unduly jeopardized the privacy of EPIC's members.

---

[1] *See, e.g.*, *Census: Australian Bureau of Statistics Says Website Attacked by Overseas Hackers*, ABC News (Aug. 9, 2016), https://www.abc.net.au/news/2016-08-10/australian-bureau-of-statistics-says-census-website-hacked/7712216 ("Australian Privacy Commissioner Timothy Pilgrim said he would launch an investigation into the ABS cyber attacks. 'My first priority is to ensure that no personal information has been compromised as a result of these attacks,' he said. 'My office will continue to work with the ABS to ensure they are taking appropriate steps to protect the personal information collected through the census.'").

EPIC has repeatedly warned the Census Bureau of the defective nature of the proposed census collection, yet the agency continues to move forward with this unlawful agency action. Key deadlines are fast approaching, and major privacy risks have not been addressed by the agency. And nowhere is the duty to assess privacy risks more important than the decennial census, a "unique" and compulsory collection of data that "reaches every population group, from America's long-time residents to its most recent immigrants." Presidential Proclamation No. 7,286, 65 Fed. Reg. 17,985, 17,985 (Apr. 1, 2000).

The Court should grant EPIC's motion for a preliminary injunction and halt the Census Bureau's implementation of the citizenship question until this Court has adjudicated EPIC's E-Government Act claims on the merits.

BACKGROUND

I. The E-Government Act of 2002

In 2002, Congress passed the E-Government Act with the aim of "provid[ing] enhanced access to Government information" and "mak[ing] the Federal Government more transparent and accountable." E-Government Act, Pub. L. No. 107-347, §§ 2(b)(9), (11), 116 Stat. 2899, 2901 (Dec. 17, 2002) (codified at 44 U.S.C. § 3501 note), *see also* 148 Cong. Rec. 11,227 (2002) (statement of Sen. Lieberman) (explaining that the Act is intended to "improv[e] the access of all citizens to the government services and information they rely on every day in their work and personal lives"). Among the "constituencies" accounted for in the Act are "the public access community," "privacy advocates," and "non-profit groups interested in good government." 148 Cong. Rec. 11,228.

Section 208 of the Act requires federal agencies to conduct and publish a privacy impact assessment ("PIA") before beginning the process of collecting personal data. E-Government Act § 208(b). Specifically, "before . . . initiating a new collection" of "information in an identifiable

form" from ten or more persons, the agency must "conduct a privacy impact assessment" and, "if practicable," "make the privacy impact assessment publicly available through the website of the agency, publication in the Federal Register, or other means." *Id.* § 208(b). Information is in an identifiable form if it "permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means." *Id.* § 208(d). To satisfy section 208 of the E-Government Act, a privacy impact assessment must specify:

> (I) what information is to be collected;
> (II) why the information is being collected;
> (III) the intended use of the agency of the information;
> (IV) with whom the information will be shared;
> (V) what notice or opportunities for consent would be provided to individuals regarding what information is collected and how that information is shared;
> (VI) how the information will be secured; and
> (VII) whether a system of records is being created under section 552a of title 5, United States Code, (commonly referred to as the 'Privacy Act').

E-Government Act § 208(b)(2)(B)(ii). A PIA must also be "commensurate with the size of the information system being assessed, the sensitivity of information that is in an identifiable form in that system, and the risk of harm from unauthorized release of that information[.]" *Id.* § 208(b)(2)(B)(i).

II. The Department of Commerce order initiating the collection of personal data concerning citizenship status

In order to determine the apportionment of representatives "among the several States," the Census Clause of the U.S. Constitution, as amended, requires that an "actual Enumeration" of persons be undertaken every ten years "in such Manner as [Congress] shall by Law direct." U.S. Const. art. 1, § 2, cl. 3; *see also* U.S. Const. amend. XIV, § 2 ("Representatives shall be apportioned among the several States according to their respective numbers, counting the whole number of persons in each State, excluding Indians not taxed."). In furtherance of the Census Clause, Congress has directed the Secretary of Commerce to "take a decennial census of

population," 13 U.S.C. § 141(a), and to "determine the inquiries, and the number, form, and subdivisions" of the questionnaires to be used in the census. 13 U.S.C. § 5. Congress has also established the Census Bureau as an agency under the Department of Commerce. 15 U.S.C. § 1501. The Bureau will administer the next Census in 2020. *See* U.S. Census Bureau, *2020 Census* (Oct. 19, 2018).[2] By law, any person who refuses to answer "any of the questions . . . submitted to him in connection with any census"—or who willfully gives a false answer to a census question—is subject to criminal penalties. 13 U.S.C. § 221(a)–(b).

On March 26, 2018, Secretary of Commerce Wilbur Ross stated that he "ha[d] determined that reinstatement of a citizenship question on the 2020 decennial census [wa]s necessary" and that he was "directing the Census Bureau to place the citizenship question last on the decennial census form." Ex. 1, Letter from Wilbur Ross, Secretary of Commerce, to Karen Dunn Kelley, Under Secretary for Economic Affairs, at 8 (Mar. 26, 2018). No such question appeared on the 2010 Census, Ex. 16, U.S. Dep't of Commerce, *United States Census 2010* (2009), nor has the Bureau posed a citizenship question to all census respondents since the 1950 Census. *See* Ex. 1 at 2. On March 28, 2018, the Bureau officially reported to Congress its intention to add a citizenship question to the 2020 Census. Ex. 2, U.S. Census Bureau, *Questions Planned for the 2020 Census and American Community Survey* 7 (March 2018). The version of the question presented to Congress asked: "Is this person a citizen of the United States?" *Id.* Five responses were listed: "Yes, born in the United States"; "Yes, born in Puerto Rico, Guam, the U.S. Virgin Islands, or Northern Marianas"; "Yes, born abroad of U.S. citizen parent or parents"; "Yes, U.S. citizen by naturalization – *Print year of naturalization*"; "No, not a U.S. Citizen[.]" *Id.*

[2] https://www.census.gov/programs-surveys/decennial-census/2020-census.html.

Secretary Ross stated that the citizenship question was added in response to a December 2017 request by the Department of Justice ("DOJ"), which he claimed sought citizenship data to enable "more effective enforcement" of the Voting Rights Act. Ex. 1 at 1. The DOJ's request raised alarm and opposition from members of Congress, the attorneys general of at least twenty states, and mayors across the country. *See* Letter from Sen. Dianne Feinstein et al. to Wilbur Ross, Secretary of Commerce (Jan. 5, 2018);[3] Letter from Attorneys General of Twenty U.S. States to Wilbur Ross, Secretary of Commerce (Feb. 12, 2018);[4] U.S. Conference of Mayors, *Nation's Mayors to Secretary Ross: Don't Politicize Census. Remove the Citizenship Question* (Mar. 27, 2018).[5] Moreover, Secretary Ross's explanation for his decision is at odds with the extensive evidence adduced in *New York v. U.S. Dep't of Commerce*, No. 18-CV-2921 (JMF), 2019 WL 190285 (S.D.N.Y. Jan. 15, 2019), including the Secretary's subsequent admission that he communicated with Chief White House strategist Steve Bannon and Kansas Secretary of State Kris Kobach about the citizenship question months before the DOJ made a request. *See, e.g.*, Defs.' Second Suppl. Resps. to Pls.' First Set of Interrogatories 2–3, *N.Y. Immigration Coal. v. U.S. Dep't of Commerce,* 18-5025 (Oct. 11, 2018);[6] Email from Kris Kobach, Sec'y, Kan. Dep't of State, to Wilbur Ross, Sec'y, Dep't of Commerce (Jul. 21, 2017).[7]

[3] https://www.feinstein.senate.gov/public/_cache/files/3/7/376f8dcd-7f35-4913-9e80-cd1e48e3b312/7E4C59B2988E2CC14866543EDD7E01A6.2018.01.05-census-citizeship-letter.pdf.
[4] https://www.brennancenter.org/sites/default/files/legal-work/Multi-State-Attorney-General-Letter-re-2020-Census.pdf.
[5] https://www.usmayors.org/2018/03/27/nations-mayors-to-secretary-ross-dont-politicize-census-remove-the-citizenship-question/.
[6] https://ag.ny.gov/sites/default/files/second_supp_res_to_rog_1_final_2018.10.11.pdf.
[7] https://epic.org/foia/censusbureau/EPIC-18-03-22-Census-Bureau-FOIA-20180611-Production-Kobach-Emails.pdf.

III. The privacy implications of collecting citizenship status information

Although the Census Bureau's collection of personally identifiable information carries inherent privacy risks, the addition of a citizenship question to the 2020 Census poses a unique threat to privacy, personal security, and the accuracy of the United States Census. The citizenship question would compel the release of respondents' citizenship status (and potentially immigration status), which could in turn expose individuals and their family members to investigation, sanction, and deportation. Secretary Ross's stated basis for adding the citizenship question was to provide the DOJ with "census block level citizen voting age population ('CVAP') data," Ex. 1 at 1, 8—data that is susceptible to reidentification. Ex. 3, Latanya Sweeney, *Simple Demographics Often Identify People Uniquely* 2 (Carnegie Mellon Univ., Data Privacy Working Paper No. 3, 2000). The Bureau has indicated that census response data—including individuals' citizenship status information—may be transferred in "[b]ulk" to other federal agencies "[f]or criminal law enforcement activities." Ex. 6, U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN08 Decennial Information Technology Division (DITD)* at 5, 7, 9 (approved Sep. 28, 2018). And in a June 12, 2018 email exchange between DOJ officials, disclosed in the course of litigation against Secretary Ross, DOJ officials "privately discussed the possibility that in the future census information could be shared with law enforcement." Tara Bahrampour, *Trump Administration Officials Suggested Sharing Census Responses with Law Enforcement, Court Documents Show*, Wash. Post (Nov. 19, 2018);[8] *see also* Decl. of Andrew Case in Supp. of Pls.' Opp'n to Defs.' Mot. Summ. J. at Ex. B, *San Jose v. Ross*, 18-2279 (N.D. Cal. Filed Nov. 16, 2018).

[8] https://www.washingtonpost.com/local/social-issues/trump-administration-officials-suggested-sharing-census-responses-with-law-enforcement-court-documents-show/2018/11/19/41679018-ec46-11e8-8679-934a2b33be52_story.html.

Historically, the misuse of census data has caused considerable harm to certain populations. For example, the 1910 census law prohibited the use of information supplied by businesses for non-statistical, non-census purposes, but there was no such prohibition regarding individual citizen data. Act of Jul. 2, 1909 (to provide for the expenses of the Thirteenth December Census, and for other purposes), ch. 2, § 25, 36 Stat. 1, 9. As a result, the Census Bureau was able to disclose census records to the Department of Justice and local draft boards during World War I. Margo Anderson & William Seltzer, *Challenges to the Confidentiality of U.S. Federal Statistics*, 1910-1965, 23 J Official Stat. 1, 6–7 (2007). Similarly, in 1920, the Department of Justice requested census data about individuals' citizenship for use in deportation cases. *Id.* at 8–9. In 1930, Congress passed a census statute prohibiting the Bureau from publishing any data identifying individuals. Act of June 18, 1929 (to provide for the fifteenth and subsequent decennial censuses and to provide for apportionment of Representatives in Congress), ch. 28, § 11, 46 Stat. 21, 25. Yet the Second War Powers Act weakened this restriction and permitted the Bureau in 1943 to provide the U.S. Secret Service with the names, addresses, occupations, and citizenship status of every Japanese American residing in the Washington, D.C. area. Margo Anderson & William Seltzer, *Census Confidentiality Under the Second War Powers Act (1942- 1947)* at 16 (Mar. 29-31, 2007) (unpublished manuscript). The Bureau also provided the War Department with census-block level data on Japanese Americans residing in western states to facilitate their internment. Comm'n on Wartime Relocation and Internment of Civilians, *Personal Justice Denied* 104-05 (1982).[9]

In 2004, an EPIC Freedom of Information Act ("FOIA") request revealed that the Census Bureau had provided the Department of Homeland Security with a list of cities containing more

[9] https://www.archives.gov/files/research/japanese-americans/justice-denied/chapter-3.pdf.

than 1,000 Arab-American residents and a zip-code level breakdown of Arab-American populations throughout the United States, sorted by country of origin. EPIC, *Department of Homeland Security Obtained Data on Arab Americans From Census Bureau* (2004);[10] *see also* Lynette Clemetson, *Homeland Security Given Data on Arab-Americans*, N.Y. Times (Jul. 30, 2004).[11] The Census Bureau and Customs and Border Protection revised their data request policies following EPIC's FOIA case. *See* Lynette Clemetson, *Census Policy On Providing Sensitive Data Is Revised*, N.Y. Times, (Aug. 31, 2004);[12] U.S. Customs and Border Protection, *Policy for Requesting Information of a Sensitive Nature from the Census Bureau* (Aug. 9, 2004).[13]

Many Americans remain fearful that their census responses will be "used against them or their loved ones" by federal agencies. *New York*, No. 18-CV-2921 (JMF), 2019 WL 190285, at *9. As the Census Bureau has previously acknowledged, "[q]uestions as to citizenship are particularly sensitive in minority communities and would inevitably trigger hostility, resentment and refusal to cooperate." *Fed'n for Am. Immigration Reform v. Klutznick*, 486 F. Supp. 564, 568 (D.D.C. 1980). This has consequences for the accuracy of census data. Fear over the misuse of census responses can lead individuals to provide false or incomplete information, which "harms the quality of the census count, and would [result in] substantially less accurate citizenship status data than are available from administrative sources." Ex. 4, Memorandum from John M. Abowd, Chief Scientist, U.S. Census Bureau, to Wilbur L. Ross, Sec'y of Commerce, at 1 (Jan. 19, 2018); *see also Fed'n for Am. Immigration Reform*, 486 F. Supp. at 568 ("[A]ccording to the

[10] https://epic.org/privacy/census/foia/.
[11] https://www.nytimes.com/2004/07/30/us/homeland-security-given-data-on-arab-americans.html.
[12] https://www.nytimes.com/2004/08/31/us/census-policy-on-providing-sensitive-data-is-revised.html.
[13] https://epic.org/privacy/census/foia/policy.pdf.

Bureau; any effort to ascertain citizenship will inevitably jeopardize the overall accuracy of the population count.").

IV. The lack of privacy impact assessments analyzing the citizenship question

None of the privacy impact assessments published by the Census Bureau address the question that Secretary Ross chose to add to the 2020 Census on March 26, 2018. When the Department of Commerce and the Census Bureau produce a PIA concerning Bureau activities, the assessment is published on the Bureau's PIA webpage. Dep't of Commerce, Office of Privacy & Open Gov't, *U.S. Census Bureau Privacy Impact Assessments (PIAs) and Privacy Threshold Analysis (PTA)* (Oct. 1, 2018).[14] The webpage lists Census Bureau systems and divisions that collect, process, and store personally identifiable information. *See id.* The page also provides links to (1) the most recent privacy impact assessment for each system, and (2) the most recent privacy threshold analysis for each system. *See id.* At least five of the CEN systems and divisions identified on the Bureau's PIA webpage (CEN05, CEN08, CEN11, CEN13, and CEN18) will be used to collect, process, and/or store personally identifiable information obtained through the 2020 Census, including citizenship data. *See* Ex. 5, U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN05 Field Systems Major Application System* 1 (approved June 22, 2018); Ex. 6 at 1, 5; Ex. 7, U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN11 Demographic Census, Surveys, and Special Processing* 1, 4 (approved June 22, 2018); Ex. 8, U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN13 Center for Economic Studies (CES)* 1 (approved June 26, 2018); Ex. 9, U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN18 Enterprise Applications* 1 (approved June 26, 2018).

EPIC, on its behalf and on behalf of its members, sought out the most recent privacy impact assessment for each CEN system slated to collect, store, or process citizenship status

[14] http://www.osec.doc.gov/opog/privacy/Census-pias.html.

information. *See* Ex. 10, Declaration of Christopher Wolf ¶ 13 (Nov. 29, 2018); Ex. 11, Declaration of Bruce Schneier ¶ 13 (Nov. 29, 2018); Ex. 12, Declaration of Harry R. Lewis ¶ 13 (Dec. 3, 2018). Although a recent PIA exists for each CEN, three of the five do not mention personal data concerning citizenship status at all, while the other two include no analysis of how the collection of citizenship data would affect the privacy of census respondents. *See* Ex. 5 at 4–5 (failing to list "citizenship" among the information collected); Ex. 8 at 3–4 (same); Ex. 9 at 3–4 (same); Ex. 6 (failing to analyze the privacy implications of collecting citizenship status information); Ex. 7 (same).

EPIC, on its behalf and on behalf of its members, sought out the privacy impact assessment for "CEN05 Field Systems Major Application" by visiting the Census Bureau PIA webpage. Ex. 5; *see, e.g.,* Ex. 10 ¶ 13. CEN05 is a "major information system" that "plans, organizes, coordinates, and carries out the Bureau's field data collection program for sample surveys, special censuses, the Economic Census, and the Decennial census." Ex. 5 at 1. As such, the Bureau intends to use this system to collect personal data concerning citizenship status for the 2020 Census. *Id.* Nonetheless, the most recent PIA for CEN05 does not acknowledge that the system would handle citizenship information or analyze the privacy impact of that data collection. *See generally id.*

EPIC, on its behalf and on behalf of its members, sought out the privacy impact assessment for "CEN08 Decennial Information Technology Division (DITD)" by visiting the Census Bureau PIA webpage. Ex. 6; *see, e.g.,* Ex. 10 ¶ 13. CEN08 is a Census Bureau division and information system "consist[ing] of both general support systems and major applications," including applications that "process response data from census tests and 2020 Census operations[.]" Ex. 6 at 1. As such, the Bureau intends to use this system to collect personal data

concerning citizenship status for the 2020 Census. *See id.* Although the most recent PIA for CEN08 acknowledges the existence of the citizenship question through a single word ("Citizenship"), *id.* at 5, the PIA contains no specific analysis of the impact of collecting citizenship status information. *See generally id.*

EPIC, on its behalf and on behalf of its members, sought out the privacy impact assessment for "CEN11 Demographic Census, Surveys, and Special Processing" by visiting the Census Bureau PIA webpage. Ex. 7; *see, e.g.,* Ex. 10 ¶ 13. CEN11 is an information system "comprised of components that support the Demographic Directorate business functions" and includes "a Commercial off the Shelf (COTS) product used by Census Bureau demographic programs for data access, transformation, reporting, and statistical analysis." Ex. 7 at 1. As such, the Bureau intends to use this system to collect personal data concerning citizenship status for the 2020 Census. *See id.* Although the most recent PIA for CEN08 acknowledges the existence of the citizenship question through a single word ("Citizenship"), *id.* at 5, the PIA contains no specific analysis of the impact of collecting citizenship status information. *See generally id.*

EPIC, on its behalf and on behalf of its members, sought out the privacy impact assessment for "CEN13 Center for Economic Studies (CES)" by visiting the Census Bureau PIA webpage. Ex. 8; *see, e.g.,* Ex. 10 ¶ 13. The "data holdings [of CEN13] include census and survey data which may contain name, gender, age, date of birth etc. from across the Census Bureau[.]" Ex. 8 at 1. As such, the Bureau intends to use this system to collect personal data concerning citizenship status for the 2020 Census. *See id.* Nonetheless, the most recent PIA for CEN13 does not acknowledge that the system would handle citizenship information or analyze the privacy impact of that data collection. *See generally id.*

EPIC, on its behalf and on behalf of its members, sought out the privacy impact assessment for "CEN 18 Enterprise Applications" by visiting the Census Bureau PIA webpage. Ex. 9; *see, e.g.,* Ex. 10 ¶ 13. CEN18 is an information system "used to deliver applications to end users of the U.S. Census Bureau network." Ex. 9 at 1. The system maintains "survey and census information," including "personal names, personal addresses, personal contact information (telephone numbers, email address), business information, occupation, medical information, tax information, etc." *Id.* As such, the Bureau intends to use this system to collect personal data concerning citizenship status for the 2020 Census. *See id.* Nonetheless, the most recent PIA for CEN18 does not acknowledge that the system would handle citizenship information or analyze the privacy impact of that data collection. *See generally id.*

V. EPIC's prior warnings to the Census Bureau of E-Government Act noncompliance

Prior to this motion, EPIC notified the Census Bureau on four occasions that the Bureau was in violation of the section 208 of the E-Government Act. On August 7, 2018, EPIC filed comments with the Bureau about the proposed questions for the 2020 Census. Ex. 13, EPIC, *Comments of EPIC to the U.S. Census Bureau* (Aug. 7, 2018); *see also Proposed Information Collection; Comment Request; 2020 Census*, 83 Fed. Reg. 26,643 (June 8, 2018). EPIC stated that the Bureau had violated its PIA obligations under section 208 and that the Bureau had "not undertaken an appropriate analysis of the privacy risks of the citizenship question." Ex. 13 at 5. On October 2, 2018, EPIC published a statement to the Senate Committee on Homeland Security and Government Affairs concerning the nomination of Steven D. Dillingham to be Director of the Census. Ex. 14, Statement from EPIC to Ron Johnson, Chairman, Senate Comm. on Homeland Sec. & Gov't Affairs, et al. at 4 (Oct. 2, 2018). EPIC told the Committee (and the Bureau) that the Bureau was in violation of section 208 and "should conduct a new PIA dealing specifically with the issues raised by the citizenship question." *Id.* at 4. On Oct. 29, 2018, EPIC

filed an amicus brief in *New York*, No. 18-CV-2921 (JMF), 2019 WL 190285, a case challenging the addition of the citizenship question to the 2020 Census. EPIC explained the Bureau's PIA obligations in detail and warned that "the Bureau has not adequately justified the collection of citizenship information or shown that it has implemented the safeguards necessary to protect the data that it collects." Br. of EPIC as Amicus Curiae in Supp. Pls.' Position at Trial at 20, *New York*, No. 18-CV-2921 (JMF), 2019 WL 190285 (S.D.N.Y. Jan. 15, 2019). Finally, on Nov. 20, 2018, EPIC filed a (and subsequently served) a complaint in this case alleging that the agency had "failed to conduct any of the privacy analysis required by the E-Government Act for a major collection of personally identifiable information." Complaint ¶ 63 (Nov. 2, 2018), ECF No. 1.

To date, the Census Bureau has not updated the PIAs for CEN05, CEN08, CEN11, CEN13, or CEN18 to include an analysis of how a citizenship question would impact privacy. The Department of Commerce recently reiterated, in a filing with the Office of Management and Budget ("OMB"), that the Bureau intends to collect citizenship status information through the 2020 Census. *See* Submission for OMB Review, 83 Fed. Reg. 67,213, 67,216 (Dec. 28, 2018). The Bureau has also stated that the "printing, addressing, and mailing of Internet invitations, reminder cards or letters, and paper questionnaire packages" for the 2020 Census will begin by June 2019. U.S. Census Bureau, *2020 Census Operational Plan: A New Design for the 21st Century* at 89 (Sept. 2017).[15]

VI. EPIC's lawsuit under the E-Government Act

On November 20, 2018, EPIC filed the instant suit against the U.S. Census Bureau and the Department of Commerce (collectively, "the Defendants," "the Census Bureau," or "the Bureau"). *See* Complaint. EPIC alleged that the Defendants violated the E-Government Act and

[15] https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-oper-plan3.pdf.

Administrative Procedure Act ("APA") in two respects. First, EPIC alleged that the Defendants took unlawful action in violation of 5 U.S.C. § 706(2) and E-Government Act § 208(b) by initiating a new collection of information without first producing required privacy impact assessments. Complaint ¶¶ 64–70 (Count I). Second, EPIC alleged that the Defendants unlawfully withheld production of required privacy impact assessments in violation of 5 U.S.C. § 706(1) and E-Government Act § 208(b). Complaint ¶¶ 71–76 (Count II). EPIC also brought a claim under the Declaratory Judgment Act, 28 U.S.C. § 2201(a). Complaint ¶¶ 77–78 (Count III). As relief, EPIC sought, *inter alia*, the suspension and revocation of the citizenship question pending completion and publication of the required privacy impact assessments. Complaint pp. 26–27.

VII. The ruling of the court in *New York v. U.S. Department of Commerce*

On January 15, 2019, the U.S. District Court for the Southern District of New York entered a final judgment in *New York v. U.S. Department of Commerce*, No. 18-CV-2921 (JMF), 2019 WL 190285, a separate lawsuit challenging the addition of the citizenship question under the APA, the Enumeration Clause of the U.S. Constitution, and the Fifth Amendment of the U.S. Constitution. The court, finding "no dispute" that Secretary Ross's decision to add a citizenship question constituted "final agency action," concluded that the Secretary "violated the APA in several respects" and "violated the public trust." *Id.* at *4, *89. Accordingly, the court "vacat[ed] Secretary Ross's decision to add a citizenship question to the 2020 census questionnaire" and "enjoin[ed] Defendants from implementing Secretary Ross's March, 26, 2018 decision or from adding a question to the 2020 census questionnaire without curing the legal defects identified" by the Court. *Id.* at *125. However, the Defendants have appealed the court's judgment to the Second Circuit, Notice of Appeal, *New York*, No. 18-CV-2921 (JMF) (filed Jan. 17, 2019), and there remains pending litigation in the Supreme Court concerning a possible deposition of

Secretary Ross in the same case. *In re Dep't of Commerce*, No. 18-557, 2018 WL 5458822, at *1 (U.S. Nov. 16, 2018) (granting Defendants' mandamus petition to review a discovery order).

ARGUMENT

The Court should grant EPIC's motion for a preliminary injunction to preserve the status quo and to allow a full adjudication of EPIC's claims on the merits. EPIC has shown that it is entitled to such an injunction. In making the final decision to add a citizenship question to the 2020 Census, the Census Bureau initiated a new collection of personal information and triggered its obligation under the E-Government Act to conduct and publish new privacy impact assessments. Yet the Bureau has failed to publish the required assessments. As a result, EPIC and its members face irreparable harm: EPIC cannot evaluate the privacy impact of the citizenship question or scrutinize any privacy protections that might be put in place for the collection of citizenship status information. The privacy of EPIC's members is also unduly threatened by the Bureau's actions, which violate key privacy safeguards established by Congress. Finally, EPIC has Article III standing to bring this case because its members have suffered concrete and particularized injuries redressable by the Court.

I. EPIC IS ENTITLED TO A PRELIMINARY INJUNCTION.

EPIC is entitled to a preliminary injunction because the Census Bureau's failure to publish legally required privacy impact assessments causes EPIC irreparable harm and conflicts with the public interest. A plaintiff is entitled to a preliminary injunction upon "a clear showing" that "(1) [the plaintiff] 'is likely to succeed on the merits'; (2) [the plaintiff] 'is likely to suffer irreparable harm in the absence of preliminary relief'; (3) the 'balance of equities' tips in [the plaintiff's] favor; and (4) 'an injunction is in the public interest.'" *Navajo Nation v. Azar*, 292 F. Supp. 3d 508, 512 (D.D.C. 2018) (quoting *Winter v. NRDC*, 555 U.S. 7, 22 (2008)).

EPIC readily satisfies each of these factors. Accordingly, the Court should enjoin the Defendants from taking any steps to collect citizenship information via the 2020 Census in order to "preserve the relative positions of the parties until a trial on the merits can be held." *Benisek v. Lamone*, 138 S. Ct. 1942, 1945 (2018) (quoting *University of Tex. v. Camenisch*, 451 U.S. 390, 395 (1981)).

A. EPIC is likely to succeed on the merits of its claims.

EPIC is likely to prevail on its claims that the Census Bureau violated the E-Government Act and Administrative Procedure Act by failing to conduct and publish privacy impact assessments for the nationwide collection of citizenship status information. Under section 208 of the E-Government Act, federal agencies must "conduct," "ensure the review of," and "make publicly available" a detailed privacy impact assessment "*before* . . . initiating a new collection of information that—"

> (I) will be collected, maintained, or disseminated using information technology; and
> (II) includes any information in an identifiable form permitting the physical or online contacting of a specific individual if identical questions have been posed to, or identical reporting requirements imposed on, 10 or more persons, other than agencies, instrumentalities, or employees of the Federal Government.

E-Government Act § 208(b)(1) (emphasis added). Agencies must also "update[]" existing PIAs "where a system change creates new privacy risks." Ex. 15, Joshua B. Bolten, Dir., OMB, Executive Office of the President, M03-22, Memorandum for Heads of Executive Departments and Agencies, Attachment A § II.B.2 (Sept. 26, 2003) ("OMB Guidance").

The Census Bureau has failed to satisfy these PIA obligations with respect to each of the five Bureau systems that would contain or be used to disseminate citizenship status information.

1. The Census Bureau is required to publish privacy impact assessments that fully address the privacy risks of collecting citizenship status information.

The Census Bureau's collection of citizenship status information is a "new collection" of personally identifiable information ("PII") that triggers the agencies' PIA obligations. E-Government Act § 208(b)(1). An individual's citizenship status constitutes PII because it "permit[s] the identity of [the] individual to whom the information applies to be reasonably inferred," E-Government Act § 208(d), when collected in "combination [with] gender, race, birth date, geographic indicator, and other descriptors" demanded on the census. Ex. 15, OMB Guidance § II.A.2. The Bureau concedes as much. *See, e.g.*, Ex. 6 at 5 (listing "Citizenship" data among the "General Personal Data" handled by the CEN08 system).

The Census Bureau's collection of personal data concerning citizenship status is also "new." The Bureau did not collect citizenship status information through the 2010 Census, *see* Ex. 16, which is the only other decennial census conducted since the E-Government Act was enacted in 2002. Secretary Ross himself acknowledged that the Bureau was undertaking a "reinstatement of a citizenship question" that had not been asked on "census surveys of the entire United States population" since 1950. Ex. 1 at 2, 8. Finally, in deciding to collect citizenship status information from every person in the United States, the Bureau has imposed "identical reporting requirements" on "10 or more persons." E-Government Act § 208(b)(1)(A)(ii)(II); *see also* 13 U.S.C. § 221(a)–(b) (requiring all persons over 18 to respond to census questions). The Bureau is thus required to conduct and publish a fully compliant PIA for each IT system that it will deploy to collect, maintain, or disseminate citizenship status information. *See* E-Government Act § 208(b)(1); Ex. 5 at 1; Ex. 6 at 1, 5; Ex. 7 at 1, 4; Ex. 8 at 1; Ex. 9 at 1.

Similarly, the Census Bureau's order initiating the collection of citizenship status information triggers the agencies' obligation to update and republish existing PIAs because "new

information in identifiable form [is being] added to a collection" that "raises the risks to personal privacy[.]" Ex. 15, OMB Guidance § II.B.2. On March 26, 2018—the day that the Bureau initiated a new collection of citizenship status information—each of the five IT systems slated to collect, maintain, or disseminate citizenship information was covered by a prior PIA. *See* Ex. 17, U.S. Dep't of Commerce, *U.S. Census Bureau Privacy Impact Assessments (PIAs) and Privacy Threshold Analysis (PTA)* (Dec. 13, 2016). But none of those prior PIAs addressed the extraordinary privacy implications of collecting and processing citizenship status information from every person in the United States, as now required by the Secretary of Commerce. The Bureau was simply not engaged in such data collection prior to the March 26, 2018 order. By initiating the addition of "new information in identifiable form" to multiple Census Bureau IT systems, the Bureau therefore triggered its obligation to fully update the existing PIAs for those systems. Ex. 15, OMB Guidance §§ II.B.2.i.

Indeed, the "new privacy risks" of adding a citizenship question to the 2020 Census are profound. Ex. 15, OMB Guidance § II.B.2. Citizenship status information is every bit as sensitive as "health or financial information," and the collection of citizenship data presents unique threats to privacy and personal security. *Id.* § II.B.2.i. Disclosure of citizenship status to law enforcement agencies or the public at large could potentially expose individuals and their family members to investigation, sanction, retaliation, and deportation. Secretary Ross has expressly stated that the Census Bureau will transfer to the Department of Justice "census block level citizenship voting age population ('CVAP') data," Ex. 1 at 1, 8—data that is susceptible to reidentification. Ex. 3 at 2. The Bureau has also indicated that individuals' citizenship status information may be transferred in "[b]ulk" to other federal agencies "[f]or criminal law enforcement activities." Ex. 6 at 5, 7, 9. These planned transfers of personal data are "significant

new uses [and] exchanges of information in an identifiable form" and constitute a "merg[ing], centraliz[ing], [or] match[ing]" of citizenship data with PII in "other databases[.]" Ex. 15, OMB Guidance §§ II.B.2.d, g.

Given the novel privacy risks posed by the collection and dissemination of citizenship status information, *id.* § II.B.2, the Census Bureau is required to fully update and republish the PIA for each IT system that will handle personal data concerning citizenship status. *See id.*; Ex. 5 at 1; Ex. 6 at 1, 5; Ex. 7 at 1, 4; Ex. 8 at 1; Ex. 9 at 1.

2. The Census Bureau was obligated to publish the required privacy impact assessments by March 26, 2018.

The deadline for the Census Bureau to conduct and publish the required privacy impact assessments was March 26, 2018. That is the day Secretary Ross ordered the addition of a citizenship question to the 2020 Census, thereby initiating a new collection of citizenship status information.

An agency must conduct, review, and publish a valid privacy impact assessment before "initiating a new collection of information[.]" E-Government Act § 208(b)(1)(A). The word "initiate" means "[c]ommence, start; originate; introduce; inchoate." *Kelso v. U.S. Dep't of State*, 13 F. Supp. 2d 1, 9 (D.D.C. 1998) (quoting Black's Law Dictionary 705 (Special Deluxe 5th ed. 1979)). Title 44, in which the E-Government Act is codified, specifically distinguishes between "initiating" a process (*i.e.*, commencing or introducing it) and "carrying out" or "completing" a process. 44 U.S.C. § 3902(a) ("The Director of the Government Publishing Office shall have no authority to prevent or prohibit the Inspector General from initiating, carrying out, or completing any audit or investigation[.]"). The phrase "collection of information" means

> the obtaining, causing to be obtained, soliciting, or requiring the disclosure to third parties or the public, of facts or opinions by or for an agency, regardless of form or format, calling for either—

> (i) answers to identical questions posed to, or identical reporting or recordkeeping requirements imposed on, ten or more persons, other than agencies, instrumentalities, or employees of the United States; or
> (ii) answers to questions posed to agencies, instrumentalities, or employees of the United States which are to be used for general statistical purposes[.]

44 U.S.C. § 3502(3)(A); *see also* E-Government Act § 201 ("Except as otherwise provided, in this title the definitions under sections 3502 and 3601 of title 44, United States Code, shall apply.").

An agency need not actually acquire "facts or opinions" for the collection of information to exist, so long as the process of "obtaining," "soliciting," or "requiring . . . disclosure" is underway. 44 U.S.C. § 3502(3)(A); *see also* 44 U.SC. § 3506(c) (requiring agencies to "review each collection of information" well before any information has been solicited or added to the collection). The OMB further defines "collection of information" to include "a *plan* and/or an instrument calling for the collection or disclosure of information[.]" 5 C.F.R. § 1320.3(c) (emphasis added); *see also EPIC v. Presidential Advisory Comm'n on Election Integrity*, 878 F.3d 371, 380 (D.C. Cir. 2017) (noting that an agency must "prepare a privacy impact assessment [if] it *plans* to collect information" (emphasis added)). Even an "algorithm" can constitute a "collection of information" under 44 U.S.C. § 3502(3)(A) if it "impose[s] a 'reporting requirement'" on members of the public. *Benkelman Tel. Co. v. FCC*, 220 F.3d 601, 607 (D.C. Cir. 2000) (quoting *Saco River Cellular, Inc. v. FCC*, 133 F.3d 25, 33 (D.C. Cir. 1998)).

Accordingly, the Census Bureau initiated a new collection of information when—on March 26, 2018—Secretary Ross ordered the "reinstatement of a citizenship question on the 2020 decennial census." Ex. 1 at 8. With the Secretary's announcement "directing the Census Bureau to place the citizenship question . . . on the decennial census form," *id.*, the Bureau both

commenced the "soliciting" and "obtaining" of citizenship status information, 44 U.S.C. § 3502(3)(A), and introduced a definite "plan . . . calling for the collection or disclosure of information," 5 C.F.R. § 1320.3(c) (construing § 3502(3)(A)). By the same token, the Bureau introduced a new requirement that census respondents disclose "facts or opinions" to "third parties or the public." 44 U.S.C. § 3502(3)(A). Persons over 18 must respond to a citizenship question on the census, 13 U.S.C. § 221(a)–(b), and the aggregate facts provided by census respondents are disclosed to the public and third parties. *See* U.S. Census Bureau, *Decennial Census of Population and Housing* (2014)[16] (publishing aggregate responses to 2010 Census questions); Ex. 1 at 8 (stating that the citizenship question "is necessary to provide complete and accurate data in response to the DOJ request" for citizenship data); U.S. Census Bureau, *Subjects Planned for the 2020 Census and American Community Survey* 52 (Mar. 2017)[17] (listing six outside agencies that would use census-collected citizenship data); Submission for OMB Review, 83 Fed. Reg. at 67,221 (promising that if state officials "indicate a need for tabulations of citizenship data . . . the Census Bureau will make a design change to include citizenship as part of that data").

Because the Census Bureau initiated a new collection of information of citizenship status information on March 26, 2018, the Bureau was obligated to conduct and publish the required privacy impact assessments by that date.

3. The Census Bureau has unlawfully failed to publish the required privacy impact assessments for more than nine months.

The Census Bureau failed to produce valid privacy impact assessments by March 26, 2018, and still refuses to correct its unlawful conduct despite EPIC's repeated warnings and the

[16] https://www.census.gov/programs-surveys/decennial-census/data/datasets.2010.html.
[17] https://www2.census.gov/library/publications/decennial/2020/operations/planned-subjects-2020-acs.pdf.

initiation of this lawsuit. To satisfy section 208 of the E-Government Act, an agency must publish a PIA specifying, *inter alia*, "what information is to be collected"; "why the information is being collected"; "the intended use of the agency of the information"; "with whom the information will be shared"; and "how the information will be secured[.]" E-Government Act § 208(b)(2)(B)(ii). A PIA also "must identify what choices the agency made regarding an IT system or collection of information as a result of performing the PIA." Ex. 15, OMB Guidance § II.C.1.b. Moreover, a PIA must be "commensurate with the size of the information system being assessed, the sensitivity of information that is in an identifiable form in that system, and the risk of harm from unauthorized release of that information[.]" E-Government Act § 208(b)(2)(B)(i); *see also* Ex. 15, OMB Guidance § II.C.1.b ("The depth and content of the PIA should be appropriate for the nature of the information to be collected and the size and complexity of the IT system."). Where a PIA is required for a "major information system[]"[18]—for example, a Census Bureau system used to collect and process hundreds of millions of census responses—the PIA "should reflect more extensive analyses of:"

1. the consequences of collection and flow of information,
2. the alternatives to collection and handling as designed,
3. the appropriate measures to mitigate risks identified for each alternative and,
4. the rationale for the final design choice or business process.

Ex. 15, OMB Guidance § II.C.2.a.ii. In such cases, a mere "checklist or template" will not satisfy an agency's PIA obligation. *Id.* § II.C.2.a.iii.

The Census Bureau has failed to publish a valid PIA for any of the five systems that the Bureau proposes to collect, maintain, or disseminate citizenship status information. With respect

[18] A "major information system" includes any "system or project that requires special management attention because of its: (i) importance to the agency mission, (ii) high development, operating and maintenance costs, (iii) high risk, (iv) high return, (v) significant role in the administration of an agency's programs, finances, property or other resources." Ex. 15, OMB Guidance § II.A.4.

to three systems (CEN05, CEN13, and CEN18), the Bureau's failure is total. Although these systems would be used to maintain or process response data from the 2020 Census, the most recent PIA for each system does not even acknowledge that citizenship status information is among the data "to be collected." E-Govermment Act § 208(b)(2)(B)(ii); *see also* Ex. 5 at 1, 4–6; Ex. 8 at 1, 3–4; Ex. 9 at 1, 3–4. Consequently, the Bureau has entirely failed to assess the privacy impact of maintaining and processing personal data concerning citizenship status with these systems. *See* Ex. 5; Ex. 8; Ex. 9.

The current PIAs for two other Census Bureau systems (CEN08 and CEN11) acknowledge, with a single word, that citizenship status information will be collected. *See* Ex. 6 at 5 ("Citizenship"); Ex. 7 at 4 ("Citizenship"). Yet neither PIA includes any analysis—let alone a "commensurate" analysis, § 208(b)(2)(B)(i)—of how the collection of this new information will impact privacy. For example, neither PIA explains "why [citizenship] information is being collected" via the census for the first time in 70 years or what "the intended use of the agency of [citizenship] information" is. § 208(b)(2)(B)(ii); *see also* Ex. 6; Ex. 7. Nor do the PIAs address the privacy "consequences of collection and flow of [citizenship] information"; "the alternatives to collection and handling [of citizenship information]"; "the rationale for the final design choice" to include citizenship information; or "what choices the agency made . . . as a result of performing the PIA." Ex. 15, OMB Guidance §§ II.C.1.b, a.ii; *see also* Ex. 6; Ex. 7. The CEN08 PIA even indicates—without explanation or analysis—that citizenship status and other census data could be transferred in "[b]ulk" to federal agencies "[f]or criminal law enforcement activities." Ex. 6 at 7, 9. Yet the privacy risks posed by this alarming (and illegal) possibility are nowhere addressed.

The Census Bureau was obligated to publish valid PIAs for its collection of citizenship status information by March 26, 2018. The Bureau failed to do so. After nine months and multiple filings by EPIC notifying the Bureau of its unlawful conduct, it has still failed to do so. The Bureau is therefore in violation of section 208 of the E-Government Act.

4. EPIC's suit satisfies the requirements for an action under the APA.

EPIC is entitled to bring this suit under the Administrative Procedure Act because EPIC has been "adversely affected or aggrieved" by the Census Bureau's violations of the E-Government Act, 5 U.S.C. § 702, which this Court has the power to review. 5 U.S.C. §§ 704, 706. Because EPIC is a membership organization, Ex. 18, *Bylaws of the Electronic Privacy Information Center* §§ 2.02, 5.01–5.05 (amended Jan. 26, 2018), it is aggrieved to the same extent that EPIC's members are. *See Pharm. Research & Mfrs. of Am. v. Thompson*, 259 F. Supp. 2d 39, 56 (D.D.C. 2003), *aff'd*, 362 F.3d 817 (D.C. Cir. 2004) (finding membership organization that "represent[ed] the interests of urban Indian Centers" therefore "ha[d] prudential standing to assert" an APA claim on behalf of those centers).

First, EPIC is aggrieved by the Census Bureau's unlawful initiation of a new collection of information without first completing the required privacy impact assessments. *Id.*; *see also* Complaint ¶¶ 64–70. By "direct[ing] the Census Bureau to place the citizenship question . . . on the decennial census form" after "a thorough review of the legal, program, and policy considerations," Ex. 1 at 8, Secretary Ross took "final agency action" on behalf of the Bureau. § 704. "There is no dispute the Secretary Ross's decision constitutes 'final agency action' reviewable under the APA." *New York*, No. 18-CV-2921 (JMF), 2019 WL 190285; *see also Kravitz v. Dep't of Commerce*, No. GJH-18-1041, 2018 WL 6830226, at *1, *6 (D. Md. Dec. 28, 2018) (analyzing the "U.S. Census Bureau's decision to include a citizenship question on the

2020 Census" as a "final agency action"). This action was final because it "mark[ed] the consummation of the agency's decisionmaking process" and is a decision "by which rights or obligations have been determined, or from which legal consequences will flow[.]" *Ipsen Biopharmaceuticals, Inc. v. Hargan*, 334 F. Supp. 3d 274, 277 (D.D.C. 2018) (quoting *U.S. Army Corps of Eng'rs v. Hawkes Co.*, 136 S. Ct. 1807, 1813 (2016)); *see also* 13 U.S.C. § 221(a)–(b) (requiring all persons over 18 to respond to census questions). Moreover, the Bureau's final agency action violated section 208 of the E-Government Act, a provision whose purpose is to "is to ensure sufficient protections for the privacy of personal information[.]" § 208(a); *see also EPIC*, 878 F.3d at 378 (explaining that section 208 of the E-Government Act "protect[s] *individuals* . . . by requiring an agency to fully consider their privacy before collecting their personal information" (emphasis in original)). EPIC is thus aggrieved by the Bureau's action, which unlawfully imperils the privacy of EPIC's members and has denied them information to which they are legally entitled. *See* Ex. 10 ¶¶ 11–15; Ex. 11 ¶¶ 11–15; Ex. 12 ¶¶ 11–15. Finally, this Court has the power to "hold unlawful and set aside" the Bureau's action because it is "not in accordance with law" and was carried out "without observance of procedure required by law[.]" § 706(2); *see also* § 704.

Second, EPIC is "aggrieved" by the Census Bureau's unlawful failure to complete the required privacy impact assessments for the collection of citizenship status information. An agency "unlawfully with[olds]" action under § 706 the APA when it fails to take a "discrete action that the agency has a duty to perform." *W. Org. of Res. Councils v. Zinke*, 892 F.3d 1234, 1241 (D.C. Cir. 2018) (citing *Norton v. Southern Utah Wilderness Alliance* (*SUWA*), 542 U.S. 55, 62–63 (2004)). Here, the Bureau failed to take the discrete actions of conducting, reviewing, and publishing required privacy impact assessments before "initiating a new collection of

information," as mandated by section § 208(b) of the E-Government Act. These duties are "nondiscretionary" and amount to "specific, unequivocal command[s]." *W. Org. of Res. Councils*, 892 F.3d at 1241 (quoting *SUWA*, 542 U.S. at 63–64); *see also* E-Government Act § 208(b)(1)(A) ("An agency shall take actions described under subparagraph (B)[.]"). By failing to conduct or publish PIAs for the collection of citizenship status information, the Bureau has failed to "ensure sufficient protections for the privacy of personal information," § 208(a), and failed to "fully consider the[ ] privacy" of individuals whose personal data would be collected. *EPIC*, 878 F.3d at 378. EPIC is thus aggrieved by the Bureau's inaction, which unlawfully imperils the privacy of EPIC's members and has denied them information to which they are legally entitled. *See* Ex. 10 ¶¶ 11–15; Ex. 11 ¶¶ 11–15; Ex. 12 ¶¶ 11–15. Finally, this Court has the power to "compel agency action unlawfully withheld or unreasonably delayed" under 5 U.S.C. § 706(1). As the D.C. Circuit recently explained, "the court can undertake review [under § 706(1)] as though the agency had denied the requested relief and can order an agency to either act or provide a reasoned explanation for its failure to act." *Friedman v. FAA*, 841 F.3d 537, 545 (D.C. Cir. 2016) (quoting *Sierra Club v. Thomas*, 828 F.2d 783, 793 (D.C. Cir. 1987)).

Because EPIC is aggrieved by the Census Bureau's violations of the E-Government Act and the APA, and because this Court has the power to review and remedy those violations, EPIC "is likely to succeed on the merits" of its claims. *Navajo Nation*, 292 F. Supp. 3d at 512 (quoting *Winter*, 555 U.S. at 22).

B. EPIC will suffer irreparable harm absent an injunction.

EPIC and its members will "suffer irreparable harm in the absence of preliminary relief" in three ways. *Navajo Nation*, 292 F. Supp. 3d at 512 (quoting *Winter*, 555 U.S. at 22). First, EPIC and its members will be irreparably harmed—and have already been harmed—by the

Census Bureau's failure to disclose "information [that] is highly relevant to an ongoing and highly public matter." *EPIC v. Presidential Advisory Comm'n on Election Integrity*, 266 F. Supp. 3d 297, 319 (D.D.C.), *aff'd on other grounds*, 878 F.3d 371. Second, EPIC and its members will be irreparably harmed—and have already been harmed—by "the failure of decision-makers" to take privacy "factors into account in the way that" the E-Government Act mandates. *Jones v. D.C. Redevelopment Land Agency*, 499 F.2d 502, 512 (D.C. Cir. 1974). Third, EPIC and its members will be irreparably harmed if the Bureau is permitted invade the privacy of EPIC's members by unlawfully compelling disclosure of their citizenship status.

First, the Census Bureau's failure to publish valid privacy impact assessments irreparably harms EPIC by denying EPIC's members information vital to the national debate over the inclusion of the citizenship question on the 2020 Census. As the D.C. Circuit has emphasized, "stale information is of little value[.]" *Judicial Watch, Inc. v. DHS*, 895 F.3d 770, 778 (D.C. Cir. 2018) (quoting *Payne Enters., Inc. v. United States*, 837 F.2d 486, 494 (D.C. Cir. 1988)); *see also Byrd v. EPA*, 174 F.3d 239, 244 (D.C. Cir. 1999) ("Byrd's injury, however, resulted from EPA's failure to furnish him with the documents until long after they would have been of any use to him."). Thus, "the non-disclosure of information to which a plaintiff is entitled, under certain circumstances itself constitutes an irreparable harm; specifically, where the information is highly relevant to an ongoing and highly public matter." *EPIC*, 266 F. Supp. 3d at 319; *see also EPIC v. DOJ*, 416 F. Supp. 2d 30, 41 (D.D.C. 2006) ("EPIC will also be precluded, absent a preliminary injunction, from obtaining in a timely fashion information vital to the current and ongoing debate surrounding the legality of the Administration's warrantless surveillance program."); *Washington Post v. DHS*, 459 F. Supp. 2d 61, 75 (D.D.C. 2006) ("Because the urgency with which the plaintiff makes its FOIA request is predicated on a matter of current

national debate, due to the impending election, a likelihood for irreparable harm exists if the plaintiff's FOIA request does not receive expedited treatment.").

The collection of citizenship status information through the 2020 Census is inarguably "an ongoing and highly public matter." *EPIC*, 266 F. Supp. 3d at 319. The Census Bureau has resolved to collect the citizenship status of hundreds of millions of persons nationwide, a controversial decision that has been the subject of tens of thousands of articles of interest to the public. *See "Census" and "Citizenship"*, Google News (Jan. 13, 2019)[19] (identifying more than 57,000 news articles containing the words "Census" and "Citizenship"). The Bureau's decision to initiate the collection of citizenship status information is the subject of at least six other federal lawsuits,[20] one of which was significant enough to warrant the Supreme Court's granting of a mandamus petition to review an interlocutory order. *In re Dep't of Commerce*, No. 18-557, 2018 WL 5458822, at *1. And the window for public debate is rapidly closing: the Bureau has stated that the "printing, addressing, and mailing of Internet invitations, reminder cards or letters, and paper questionnaire packages" for the 2020 Census must begin by June 2019. U.S. Census Bureau, *2020 Census Operational Plan: A New Design for the 21st Century*, *supra*, at 89; *see also New York*, No. 18-CV-2921 (JMF), 2019 WL 190285, at *88 (noting that "the Census Bureau 'need[s] to begin printing the 2020 census questionnaire' in June 2019"); Ex. 19, *Census Progress Report: Hearing Before the H. Comm. on Gov't Oversight & Reform*, 115 Cong. 9

[19] https://www.google.com/search?q=%22Census%22+and+%22Citizenship%22&tbm=nws.
[20] *See New York*, No. 18-CV-2921 (JMF), 2019 WL 190285; *New York Immigration Coal. v. Dep't of Commerce*, No. 18-CV-5025 (JMF), 2019 WL 190285 (S.D.N.Y. Jan. 15, 2019); *Kravitz v. Dep't of Commerce*, No. GJH-18-1041, 2018 WL 6830226; *La Union del Pueblo Entero v. Ross*, No. GJH-18-1570, 2018 WL 5885528 (D. Md. Nov. 9, 2018); *California v. Ross*, No. 18-1865 (N.D. Cal. filed March 26, 2018); *City of San Jose v. Ross*, No. 18-2279 (N.D. Cal. filed Apr. 17, 2018).

(2018) (statement of Ron Jarmin, Dir., U.S. Census Bureau) ("The physical printing of the majority of the paper materials for the 2020 Census will begin in June 2019.").

The privacy impact assessments that the Census Bureau has unlawfully withheld are also "highly relevant" to the ongoing debate over the planned citizenship question. *EPIC*, 266 F. Supp. 3d at 319. In failing to publish valid assessments, the Bureau has denied EPIC and its members critical details about the methods and systems used to process personal data concerning citizenship status, leaving EPIC "little if any information about prospective [] harms and potential mitigating measures." *Winter*, 555 U.S. at 23. The Bureau has thus irreparably harmed—and will continue to harm—the ability of EPIC's members to learn about, analyze, and debate the privacy implications of collecting citizenship status information through the census. *See* Ex. 10 ¶¶ 14–15 ("I am unable to determine whether the Census Bureau has fully considered or addressed the risks to my privacy, even as the Bureau begins to develop a new collection of personal data that will contain my citizenship status."); Ex. 11 ¶¶ 14–15 (same); Ex. 12 ¶¶ 14–15 (same). The requested injunction would avoid further such harm to EPIC by ensuring that the Bureau does not acquire any personal data concerning citizenship status before this Court has fully assessed the Bureau's compliance (or non-compliance) with the E-Government Act. Absent such an injunction, EPIC will suffer harm "beyond remediation." *Chaplaincy of Full Gospel Churches v. England,* 454 F.3d 290, 297 (D.C. Cir. 2006). EPIC's ability to meaningfully debate or challenge a planned act of government—here, the acquisition of citizenship information—cannot be restored after the government has already acted.

Second, the Census Bureau's failure to conduct and publish valid privacy impact assessments causes irreparable harm to EPIC and its members by undermining the "protections for the privacy of personal information" established by Congress. E-Government Act § 208(a).

The E-Government Act "protect[s] *individuals* . . . by requiring an agency to fully consider their privacy before collecting their personal information." *EPIC*, 878 F.3d at 378 (emphasis in original). When an agency bypasses the privacy safeguards set out in section 208, the agency unlawfully endangers the privacy of those whose personal data will be collected. The E-Government Act is directly analogous to National Environmental Policy Act ("NEPA") in this way. The NEPA requires federal agencies to publish an environment impact statement ("EIS") before taking "actions significantly affecting the quality of the human environment." 42 U.S.C. § 4332(2)(C). The D.C. Circuit has long held that a violation of the NEPA's EIS requirement—like a violation of the E-Government Act's PIA requirement—works an irreparable harm by undermining the agency's decisionmaking process:

> The harm against which NEPA's impact statement requirement was directed was not solely or even primarily adverse consequences to the environment; such consequences may ensue despite the fullest compliance. Rather NEPA was intended to ensure that decisions about federal actions would be made only after responsible decisionmakers had fully adverted to the environmental consequences of the actions, and had decided that the public benefits flowing from the actions outweighed their environmental costs. Thus, the harm with which courts must be concerned in NEPA cases is not, strictly speaking, harm to the environment, but rather the failure of decision-makers to take environmental factors into account in the way that NEPA mandates. And, for purposes of deciding whether equitable relief is appropriate, we think that this harm matures simultaneously with NEPA's requirements, *i.e.*, at the time the agency is, under NEPA, obliged to file the impact statement and fails to do so.

Jones, 499 F.2d at 512 (citing *Calvert Cliffs' Coordinating Comm., Inc. v. AEC*, 449 F.2d 1109, 1115 (1971); *Lathan v. Volpe*, 455 F.2d 1111, 1116–17 (9th Cir. 1971)); *accord Illinois Commerce Comm'n v. ICC*, 848 F.2d 1246, 1259 (D.C. Cir. 1988). So too here: the failure of the Bureau to take privacy risks into account when reviewing its proposed collection, as mandated by the E-Government Act, works an irreparable harm to those whose citizenship status information will be collected (including EPIC's members). That harm has already "mature[d],"

in that the Bureau was "obliged to file the impact statement and fail[ed] to do so." *Jones*, 499 F.2d at 512.

Moreover, a preliminary injunction will ensure that compliance with the E-Government Act "take[s] place before there has been an 'irretrievable commitment of resources,'" *Jones*, 499 F.2d at 512 (quoting *Lathan*, 455 F.2d at 1121), such as the "printing, addressing, and mailing of Internet invitations, reminder cards or letters, and paper questionnaire packages" in June 2019. U.S. Census Bureau, *2020 Census Operational Plan: A New Design for the 21st Century*, *supra*, at 89. As the D.C. Circuit has explained regarding environmental impact statements:

> The purpose of equitable intervention in cases in which federal agencies have failed to comply with NEPA's requirements is to ensure that such compliance will take place before there has been an 'irretrievable commitment of resources.' It may be that preparation of the statement will, in the end, not move the agency from adherence to decisions it has already made. But it is not for the courts to prejudge. So long as the status quo is maintained, so long as the environmental impact statement is not merely a justification for a fait accompli, there is a possibility that the statement will lead the agency to change its plans in ways of benefit to the environment. It is this possibility that the courts should seek to preserve.

Jones, 499 F.2d at 512–13 (quoting *Lathan*, 455 F.2d at 1121); *cf. Pub. Employees for Envtl. Responsibility v. United States Fish & Wildlife Serv.*, 189 F. Supp. 3d 1, 2 (D.D.C. 2016) (quoting *Realty Income Tr. v. Eckerd*, 564 F.2d 447, 456 (D.C. Cir. 1977)) ("[W]hen an action is being undertaken in violation of NEPA, there is a presumption that injunctive relief should be granted against continuation of the action until the agency brings itself into compliance."). For the same reasons, the Court should issue a preliminary injunction here to preserve the possibility that the Census Bureau will "change its plans in ways of benefit to" privacy. *Jones*, 499 F.2d at 512–13. Without such an injunction, the Bureau's decisionmaking process—and the privacy interests of EPIC's members—will be further harmed "beyond remediation." *Chaplaincy of Full Gospel Churches,* 454 F.3d at 297.

Finally, EPIC and its members will be irreparably harmed if the Census Bureau unlawfully acquires personal data concerning citizenship status from EPIC's members. The nonconsensual or compelled disclosure of one's personal information to a party who is not legally entitled to it constitutes irreparable harm. *United States v. Hubbard*, 650 F.2d 293, 314 n.73 (D.C. Cir. 1980) (finding a "serious threat of irreparable harm to the property and the privacy interests advanced" where the government seized documents from a church containing personal information of the church's members); *see also Hirschfeld v. Stone*, 193 F.R.D. 175, 187 (S.D.N.Y. 2000) ("The harm at issue here—disclosure of confidential information—is the quintessential type of irreparable harm that cannot be compensated or undone by money damages."); *Nat'l Fed'n of Fed. Emps. v. Greenberg*, 789 F. Supp. 430, 438 (D.D.C. 1992) ("[I]rreparable injury flows not only from the employee's failure to complete the form or to complete it in a manner that the Department deems unacceptable; the intrusive nature of the compelled disclosure itself also constitutes such injury."), *vacated on other grounds*, 983 F.2d 286 (D.C. Cir. 1993).

Absent an injunction blocking the Census Bureau's acquisition of citizenship status information and printing of census forms, EPIC's members will necessarily suffer such harm. Just two weeks ago, the Bureau reiterated its earlier decision to collect citizenship status information in a filing to the OMB. Submission for OMB Review, 83 Fed. Reg. at 67,216 ("The proposed questions for the 2020 Census questionnaire include age, citizenship, Hispanic origin, race, relationship, sex, and tenure.").

The Census Bureau's collection of citizenship status information will irreparably harm EPIC by denying it information critical to an ongoing nationwide debate, by disregarding privacy safeguards established by Congress, and by unlawfully compelling disclosure of personal data

concerning citizenship status from EPIC's members. Each day that the Bureau fails to produce valid privacy impact assessments, the census printing deadline (and the census itself) approaches, and the irreparability of the harms suffered by EPIC increases. The Court should therefore enter a preliminary injunction to prevent irreparable injury to EPIC and to ensure adequate time for a final decision on the merits of EPIC's claims.

C. The equities and the public interest weigh in favor of an injunction.

The Court should enter a preliminary injunction as an equitable matter because "the 'balance of equities' tips in [the plaintiff's] favor" and "'an injunction is in the public interest.'" *Navajo Nation*, 292 F. Supp. 3d at 512 (quoting *Winter*, 555 U.S. at 22). These final two factors of the four-factor test "merge when the Government is the opposing party." *Id.* (quoting *Nken v. Holder*, 556 U.S. 418, 435 (2009)).

EPIC has a powerful equitable interest in avoiding the irreparable harms that will result if the Census Bureau continues its unlawful collection of citizenship status information without first conducting and publishing valid privacy impact assessments. *See* Part I.B, *supra*. For example, the equities favor enabling EPIC to review to the agency's assessment of privacy risks *before* the citizenship question is fully implemented. *See EPIC*, 266 F. Supp. 3d at 319 (holding that a denial of "information [that] is highly relevant to an ongoing and highly public matter" works an irreparable harm). Indeed, EPIC's informational interest in this matter is uniquely strong. EPIC is the premiere organization in the country working to enforce agencies' E-Government Act obligations, Complaint ¶ 8 (collecting EPIC PIA cases), and EPIC has been a leading advocate of census privacy protections for nearly two decades. *Id.* ¶ 9. As an organization dedicated to both privacy and open government, EPIC is uniquely equipped to evaluate the Bureau's PIAs and to inform EPIC's members and the public about the Bureau's

compliance with the E-Government Act. *See* Complaint ¶ 6; EPIC, *EPIC FOIA Cases* (Sep. 17, 2018)[21] (collecting dozens of EPIC Freedom of Information Act cases resulting in the disclosure of thousands of privacy-related records). EPIC also has a strong equity in ensuring E-Government Act compliance "before there has been an 'irretrievable commitment of resources'" by the Bureau, *Jones*, 499 F.2d at 512 (quoting *Lathan*, 455 F.2d at 1121), and EPIC's members have an urgent need to avoid the unlawful collection of their citizenship status information. *See* Ex. 10 ¶ 15; Ex. 11 ¶ 15; Ex. 12 ¶ 15.

The equities also favor an injunction because the Census Bureau has ignored four prior warnings from EPIC, including the filing of the Complaint in this matter, concerning the Bureau's noncompliance with section 208. First, on August 7, 2018, EPIC explained in comments to the Census Bureau that the Bureau had "not undertaken an appropriate analysis of the privacy risks of the citizenship question." Ex. 13 at 5. Second, on October 2, 2018, EPIC published a statement to the Senate Committee on Homeland Security & Government Affairs urging—in advance of a hearing on the Census—that the Bureau "should conduct a new PIA dealing specifically with the issues raised by the citizenship question." Ex. 14 at 4. Third, on Oct. 29, 2018, EPIC filed an amicus brief in *New York v. Department of Commerce* arguing that "the Bureau has not adequately justified the collection of citizenship information or shown that it has implemented the safeguards necessary to protect the data that it collects." Br. of EPIC as Amicus Curiae in Supp. Pls.' Position at Trial at 20, *New York*, No. 18-CV-2921 (JMF), 2019 WL 190285. Finally, on Nov. 20, 2018, EPIC filed a (and subsequently served) a Complaint in this case alerting the Defendants that they had "failed to conduct any of the privacy analysis required by the E-Government Act for a major collection of personally identifiable information."

[21] https://www.epic.org/foia/.

Complaint ¶ 63. Despite these repeated warnings from EPIC over the past five months, the Bureau has taken no apparent measures to remedy its noncompliance with section 208. Instead, the Bureau has doubled down on its collection of citizenship status information. *See* Submission for OMB Review, 83 Fed. Reg. at 67,216.

There are few, if any, equities "to overcome the much more substantial countervailing harms" to EPIC. *League of Women Voters of United States v. Newby*, 838 F.3d 1, 13 (D.C. Cir. 2016)" Although federal agencies and the public have an interest in the sound administration of federal law, "there is generally no public interest in the perpetuation of unlawful agency action." *Id.* at 12 (citing *Pursuing America's Greatness v. FEC*, 831 F.3d 500, 511-12 (D.C. Cir. 2016); *Gordon v. Holder*, 721 F.3d 638, 653 (D.C. Cir. 2013)). To the contrary: "there is a substantial public interest in having governmental agencies abide by the federal laws that govern their existence and operations." *Id*.

Neither the Census Bureau nor the public at large has a cognizable interest in the collection of citizenship status information absent the required assessments of privacy risks. Congress has already balanced the competing interests of privacy and data collection by enacting E-Government § 208(b), which prohibits collection without assessment. The Court should not hesitate to equitably enforce that Congressionally-struck balance. *Cf. Realty Income Tr.*, 564 F.2d at 456 ("[W]hen an action is being undertaken in violation of NEPA, there is a presumption that injunctive relief should be granted against continuation of the action until the agency brings itself into compliance."); *Youngstown Sheet & Tube Co. v. Sawyer*, 343 U.S. 579, 609–10 (1952) (Frankfurter, J., concurring). ("When Congress itself has struck the balance, has defined the weight to be given the competing interests, a court of equity is not justified in ignoring that pronouncement under the guise of exercising equitable discretion.").

Even if the Census Bureau were legally permitted to collect citizenship status information without first evaluating the attendant privacy risks, the public interest would strongly disfavor such a collection. Permitting federal agencies to collect personal data without first conducting a privacy impact assessment endangers privacy and curtails public understanding of government decisionmaking. That is why Congress imposed a government-wide PIA requirement in the first place: "[t]o make the Federal Government more transparent and accountable" and "to ensure sufficient protections for the privacy of personal information[.]" E-Government Act §§ 2(b)(9), 208(a). Nowhere is the duty to assess privacy risks more important than the decennial census, a "unique" and compulsory collection of data that "reaches every population group, from America's long-time residents to its most recent immigrants." Presidential Proclamation No. 7,286, 65 Fed. Reg. at 17,985. Moreover, the public interest leans heavily against the implementation of the citizenship question because—as the Census Bureau's chief scientist recently noted—adding the question would be "very costly, harms the quality of the census count, and would [result in] substantially less accurate citizenship status data than are available from administrative sources." *See* Ex. 4 at 1.

The public interest also favors injunctive relief at this stage because there is likely to be an "irretrievable"—and potentially wasteful—"commitment of resources" by the Census Bureau absent a preliminary order halting the citizenship question. *Jones*, 499 F.2d at 512 (quoting *Lathan*, 455 F.2d at 1121). As the Bureau has argued, "finalizing the decennial census questionnaire is time-sensitive," and it is "important to resolve as soon as possible" the key legal issues concerning the 2020 Census. Petition for a Writ of Mandamus at 15, *In re Dep't of Commerce*, No. 18-557, 2018 WL 5458822 (U.S. Nov. 16, 2018). If no injunction issues prior to final judgment in this case, the Bureau will likely expend significant resources in the meantime

implementing the citizenship question—a question that this Court may later deem unlawful. *See, e.g.*, U.S. Census Bureau, *2020 Census Operational Plan: A New Design for the 21st Century*, *supra*, at 89 (noting that the "printing, addressing, and mailing of Internet invitations, reminder cards or letters, and paper questionnaire packages" for the 2020 Census will begin by June 2019).

Because the balance of the equities and the public interest strongly favor a halt to the Census Bureau's implementation of the citizenship question, the Court should grant EPIC's motion for a preliminary injunction. The requested injunction is carefully tailored to address the privacy implications of the citizenship question and poses no impediment to other census operations or questions.

II. EPIC HAS ARTICLE III STANDING IN THIS MATTER.

EPIC has Article III standing to bring this suit because EPIC's members have suffered "concrete injur[ies]-in-fact that [are] fairly traceable to the defendant[s'] action[s] and capable of being redressed by a favorable judicial decision." *Conference of State Bank Supervisors v. Office of Comptroller of Currency*, 313 F. Supp. 3d 285, 294 (D.D.C. 2018) (quoting *Summers v. Earth Island Inst.*, 555 U.S. 488, 493 (2009)). For the purposes of determining Article III jurisdiction, courts must accept the plaintiff's "view of the law" on which the plaintiff's claims are based, *FEC v. Akins*, 524 U.S. 11, 21 (1998), so long as it is nonfrivolous. *See Steel Co. v. Citizens for a Better Env't*, 523 U.S. 83, 89 (1998) ("It is firmly established in our cases that the absence of a valid (as opposed to arguable) cause of action does not implicate subject-matter jurisdiction, *i.e.*, the courts' statutory or constitutional power to adjudicate the case.").

EPIC is a membership association. Ex. 18 §§ 2.02, 5.01–5.05. As such, EPIC has standing to sue on behalf of its members whenever "(1) [EPIC's] members would otherwise have standing to sue in their own right; (2) the interests [EPIC] seeks to protect are germane to the

organization's purpose; and (3) neither the claim asserted nor the relief requested requires the participation of individual members in the lawsuit." *Sorenson Commc'ns, LLC v. FCC*, 897 F.3d 214, 224 (D.C. Cir. 2018) (quoting *Ctr. for Sustainable Econ. v. Jewell*, 779 F.3d 588, 596 (D.C. Cir. 2015)) (internal quotation marks omitted). EPIC readily satisfies these criteria here.

First, EPIC's members would have standing to bring this suit on their own because they have suffered judicially redressable injuries caused by the Census Bureau that are "concrete, particularized, and actual or imminent." *Conference of State Bank Supervisors*, 313 F. Supp. 3d at 295 (quoting *Clapper v. Amnesty Int'l USA*, 568 U.S. 398, 409 (2013)). EPIC's members sought and were denied information which the Government is required to disclose under section 208 of the E-Government Act: namely, "comprehensive and meaningful" assessments of how the collection, processing, and storage of their personal citizenship information will affect their privacy. Ex. 15, OMB Guidance § II.C.2.a.b; *see also* Ex. 10 ¶¶ 11–15; Ex. 11 ¶¶ 11–15; Ex. 12 ¶¶ 11–15. These injuries are concrete: "a plaintiff suffers an 'injury in fact' when"—as here—"the plaintiff fails to obtain information which must be publicly disclosed pursuant to a statute." *Akins*, 524 U.S. at 21 (quoting *Public Citizen v. DOJ*, 491 U.S. 440, 449 (1989)); *see also Spokeo, Inc. v. Robins*, 136 S. Ct. 1540, 1549–50 (2016) (reiterating the *Akins–Public Citizen* rule that a denial of information that Congress has ordered disclosed "constitute[s] injury in fact"). These injuries are particularized: EPIC's members each have an individual interest in obtaining the information unlawfully denied to them. *See Public Citizen*, 491 U.S. at 449–50 ("The fact that other citizens or groups of citizens might make the same complaint after unsuccessfully demanding disclosure . . . does not lessen appellants' asserted injury[.]"); *EPIC*, 878 F.3d at 378 (emphasizing that section 208 of the E-Government Act "protect[s] *individuals*"). These injuries are actual: the Government was required—and failed—to disclose

the information sought by EPIC's members *before* initiating a collection of personal data concerning citizenship status. E-Government Act § 208(b)(1)(A). And these injuries would be redressed by favorable decision of this Court ordering the Census Bureau (1) to conduct and publish legally sufficient privacy impact assessments and (2) to halt the process of collecting personal citizenship information unless and until such PIAs have been completed. *See* Compl. p. 27 ("Requested Relief").

Second, the interests EPIC seeks to protect through this suit—the right to privacy and access to information about government data collection activities—are the very core of EPIC's organizational mission. *E.g.*, Ex. 18 § 1.02 ("The Corporation is organized for the charitable and educational purposes of promoting personal privacy and constitutional rights."); Ex. 10 ¶ 4 ("Central to EPIC's mission is oversight of government activities that impact individual privacy, free expression, and democratic values."); EPIC, *EPIC FOIA Cases, supra*. EPIC therefore satisfies the requirement of "pertinence between litigation subject and organizational purpose." *Ctr. for Sustainable Econ.*, 779 F.3d at 597.

Finally, neither the claims nor the relief at issue in this case require the direct participation of EPIC's members. "Member participation is not required where a 'suit raises a pure question of law' and neither the claims pursued nor the relief sought require the consideration of the individual circumstances of any aggrieved member of the organization." *Id.* at 597 (quoting *Int'l Union, United Auto., Aerospace & Agric. Implement Workers of Am. v. Brock*, 477 U.S. 274, 287–88 (1986)). The Government's unlawful initiation of a collection of information and failure to publish valid PIAs present pure questions of law. This suit does not implicate the specific factual circumstances of individual EPIC members beyond what those

members have already declared. *See* Ex. 10; Ex. 11; Ex. 12. EPIC therefore has associational standing to bring this suit on behalf of EPIC's members.

CONCLUSION

For the foregoing reasons, the Court should grant EPIC's motion for a preliminary injunction.

Respectfully Submitted,

MARC ROTENBERG, D.C. Bar #422825
EPIC President and Executive Director

ALAN BUTLER, D.C. Bar #1012128
EPIC Senior Counsel

/s/ John Davisson
JOHN DAVISSON, D.C. Bar #1531914
EPIC Counsel

ELECTRONIC PRIVACY
INFORMATION CENTER
1718 Connecticut Avenue, N.W.
Suite 200
Washington, D.C. 20009
(202) 483-1140 (telephone)
(202) 483-1248 (facsimile)

Attorneys for Plaintiff EPIC

Dated: January 18, 2018

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

ELECTRONIC PRIVACY INFORMATION CENTER, Plaintiff, v. UNITED STATES DEPARTMENT OF COMMERCE, et al., Defendants.	Civ. Action No. 18-2711 (DLF)

[PROPOSED] ORDER

Upon consideration of the Plaintiff's Motion for a Preliminary Injunction, the Defendants' Response, and the entire record, it is hereby

ORDERED that the Plaintiff's Motion is **GRANTED**; and it is

FURTHER ORDERED that Defendants United States Department of Commerce and Bureau of the Census are enjoined from implementing the Defendants' March 26, 2018 decision to add a citizenship question to the 2020 Census; and it is

FURTHER ORDERED that Defendants United States Department of Commerce and Bureau of the Census are enjoined from initiating any collection of citizenship status information that would be obtained through the 2020 Census.

Date: _____ _____

DABNEY L. FRIEDRICH
United States District Judge

EXHIBIT LIST

Letter from Wilbur Ross, Secretary of Commerce, to Karen Dunn Kelley, Under Secretary for Economic Affairs, at 8 (Mar. 26, 2018) ..... Ex. 1

U.S. Census Bureau, *Questions Planned for the 2020 Census and American Community Survey* (March 2018) ..... Ex. 2

Latanya Sweeney, *Simple Demographics Often Identify People Uniquely* (Carnegie Mellon Univ., Data Privacy Working Paper No. 3, 2000) ..... Ex. 3

Memorandum from John M. Abowd, Chief Scientist, U.S. Census Bureau, to Wilbur L. Ross, Sec'y of Commerce (Jan. 19, 2018) ..... Ex. 4

U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN05 Field Systems Major Application System* (approved June 22, 2018) ..... Ex. 5

U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN08 Decennial Information Technology Division (DITD)* (approved Sep. 28, 2018) ..... Ex. 6

U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN11 Demographic Census, Surveys, and Special Processing* ..... Ex. 7

U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN13 Center for Economic Studies (CES)* (approved June 26, 2018) ..... Ex. 8

U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN18 Enterprise Applications* (approved June 26, 2018) ..... Ex. 9

Declaration of Christopher Wolf (Nov. 29, 2018) ..... Ex. 10

Declaration of Bruce Schneier (Nov. 29, 2018) ..... Ex. 11

Declaration of Harry R. Lewis (Dec. 3, 2018) ..... Ex. 12

EPIC, *Comments of EPIC to the U.S. Census Bureau* (Aug. 7, 2018) ..... Ex. 13

Statement from EPIC to Ron Johnson, Chairman, Senate Comm. on Homeland Sec. & Gov't Affairs, et al. (Oct. 2, 2018) ..... Ex. 14

Joshua B. Bolten, Dir., OMB, Executive Office of the President, M03-22, Memorandum for Heads of Executive Departments and Agencies, Attachment A (Sept. 26, 2003) ("OMB Guidance") ..... Ex. 15

U.S. Dep't of Commerce, *United States Census 2010* (2009) ..... Ex. 16

U.S. Dep't of Commerce, *U.S. Census Bureau Privacy Impact Assessments (PIAs) and Privacy Threshold Analysis (PTA)* (Dec. 13, 2016)..... Ex. 17

Bylaws of the Electronic Privacy Information Center (amended Jan. 26, 2018)..... Ex. 18

Census Progress Report: Hearing Before the H. Comm. on Gov't Oversight & Reform, 115 Cong. 9 (2018) (statement of Ron Jarmin, Dir., U.S. Census Bureau) ..... Ex. 19

Exhibit 1

UNITED STATES DEPARTMENT OF COMMERCE
The Secretary of Commerce
Washington, D.C. 20230

To: Karen Dunn Kelley, Under Secretary for Economic Affairs

From: Secretary Wilbur Ross

Date: March 26, 2018

Re: Reinstatement of a Citizenship Question on the 2020 Decennial Census Questionnaire

Dear Under Secretary Kelley:

As you know, on December 12, 2017, the Department of Justice ("DOJ") requested that the Census Bureau reinstate a citizenship question on the decennial census to provide census block level citizenship voting age population ("CVAP") data that are not currently available from government survey data ("DOJ request"). DOJ and the courts use CVAP data for determining violations of Section 2 of the Voting Rights Act ("VRA"), and having these data at the census block level will permit more effective enforcement of the Act. Section 2 protects minority population voting rights.

Following receipt of the DOJ request, I set out to take a hard look at the request and ensure that I considered all facts and data relevant to the question so that I could make an informed decision on how to respond. To that end, the Department of Commerce ("Department") immediately initiated a comprehensive review process led by the Census Bureau.

The Department and Census Bureau's review of the DOJ request – as with all significant Census assessments – prioritized the goal of obtaining *complete and accurate data*. The decennial census is mandated in the Constitution and its data are relied on for a myriad of important government decisions, including apportionment of Congressional seats among states, enforcement of voting rights laws, and allocation of federal funds. These are foundational elements of our democracy, and it is therefore incumbent upon the Department and the Census Bureau to make every effort to provide a complete and accurate decennial census.

At my direction, the Census Bureau and the Department's Office of the Secretary began a thorough assessment that included legal, program, and policy considerations. As part of the process, I also met with Census Bureau leadership on multiple occasions to discuss their process for reviewing the DOJ request, their data analysis, my questions about accuracy and response rates, and their recommendations. At present, the Census Bureau leadership are all career civil servants. In addition, my staff and I reviewed over 50 incoming letters from stakeholders, interest groups, Members of Congress, and state and local officials regarding reinstatement of a citizenship question on the 2020 decennial census, and I personally had specific conversations on

the citizenship question with over 24 diverse, well informed and interested parties representing a broad range of views. My staff and I have also monitored press coverage of this issue.

Congress has delegated to me the authority to determine which questions should be asked on the decennial census, and I may exercise my discretion to reinstate the citizenship question on the 2020 decennial census, especially based on *DOJ*'s request for improved CVAP data to enforce the VRA. By law, the list of decennial census questions is to be submitted two years prior to the decennial census – in this case, no later than March 31, 2018.

The Department's review demonstrated that collection of citizenship data by the Census has been a long-standing historical practice. Prior decennial census surveys of the entire United States population consistently asked citizenship questions up until 1950, and Census Bureau surveys of sample populations continue to ask citizenship questions to this day. In 2000, the decennial census "long form" survey, which was distributed to one in six people in the U.S., included a question on citizenship. Following the 2000 decennial census, the "long form" sample was replaced by the American Community Survey ("ACS"), which has included a citizenship question since 2005. Therefore, the citizenship question has been well tested.

DOJ seeks to obtain CVAP data for census blocks, block groups, counties, towns, and other locations where potential Section 2 violations are alleged or suspected, and DOJ states that the current data collected under the ACS are insufficient in scope, detail, and certainty to meet its purpose under the VRA. The Census Bureau has advised me that the census-block-level citizenship data requested by DOJ are not available using the annual ACS, which as noted earlier does ask a citizenship question and is the present method used to provide DOJ and the courts with data used to enforce Section 2 of the VRA. The ACS is sent on an annual basis to a sample of approximately 2.6 percent of the population.

To provide the data requested by DOJ, the Census Bureau initially analyzed three alternatives: Option A was to continue the status quo and use ACS responses; Option B was placing the ACS citizenship question on the decennial census, which goes to every American household; and Option C was not placing a question on the decennial census and instead providing DOJ with a citizenship analysis for the entire population using federal administrative record data that Census has agreements with other agencies to access for statistical purposes.

Option A contemplates rejection of the DOJ request and represents the status quo baseline. Under Option A, the 2020 decennial census would not include the question on citizenship that DOJ requested and therefore would not provide DOJ with improved CVAP data. Additionally, the block-group level CVAP data currently obtained through the ACS has associated margins of error because the ACS is extrapolated based on sample surveys of the population. Providing more precise block-level data would require sophisticated statistical modeling, and if Option A is selected, the Census Bureau advised that it would need to deploy a team of experts to develop model-based methods that attempt to better facilitate *DOJ*'s request for more specific data. But the Census Bureau did not assert and could not confirm that such data modeling is possible for census-block-level data with a sufficient degree of accuracy. Regardless, *DOJ*'s request is based at least in part on the fact that existing ACS citizenship data sets lack specificity and

completeness. Any future modeling from these incomplete data would only compound that problem.

Option A would provide no improved citizenship count, as the existing ACS sampling would still fail to obtain *actual*, complete number counts, especially for certain lower population areas or voting districts, and there is no guarantee that data could be improved using small-area modeling methods. Therefore, I have concluded that Option A is not a suitable option.

The Census Bureau and many stakeholders expressed concern that **Option B,** which would add a citizenship question to the decennial census, would negatively impact the response rate for non-citizens. A significantly lower response rate by non-citizens could reduce the accuracy of the decennial census and increase costs for non-response follow up ("NRFU") operations. However, neither the Census Bureau nor the concerned stakeholders could document that the response rate would in fact decline materially. In discussing the question with the national survey agency Nielsen, it stated that it had added questions from the ACS on sensitive topics such as place of birth and immigration status to certain short survey forms without any appreciable decrease in response rates. Further, the former director of the Census Bureau during the last decennial census told me that, while he wished there were data to answer the question, none existed to his knowledge. Nielsen's Senior Vice President for Data Science and the former Deputy Director and Chief Operating Officer of the Census Bureau under President George W. Bush also confirmed that, to the best of their knowledge, no empirical data existed on the impact of a citizenship question on responses.

When analyzing Option B, the Census Bureau attempted to assess the impact that reinstatement of a citizenship question on the decennial census would have on response rates by drawing comparisons to ACS responses. However, such comparative analysis was challenging, as response rates generally vary between decennial censuses and other census sample surveys. For example, ACS self-response rates were 3.1 percentage points less than self-response rates for the 2010 decennial census. The Bureau attributed this difference to the greater outreach and follow-up associated with the Constitutionally-mandated decennial census. Further, the decennial census has differed significantly in nature from the sample surveys. For example, the 2000 decennial census survey contained only eight questions. Conversely, the 2000 "long form" sample survey contained over 50 questions, and the Census Bureau estimated it took an average of over 30 minutes to complete. ACS surveys include over 45 questions on numerous topics, including the number of hours worked, income information, and housing characteristics.

The Census Bureau determined that, for 2013-2016 ACS surveys, nonresponses to the citizenship question for non-Hispanic whites ranged from 6.0 to 6.3 percent, for non-Hispanic blacks ranged from 12.0 to 12.6 percent, and for Hispanics ranged from 11.6 to 12.3 percent. However, these rates were comparable to nonresponse rates for other questions on the 2013 and 2016 ACS. Census Bureau estimates showed similar nonresponse rate ranges occurred for questions on the ACS asking the number times the respondent was married, 4.7 to 6.9 percent; educational attainment, 5.6 to 8.5 percent; monthly gas costs, 9.6 to 9.9 percent; weeks worked in the past 12 months, 6.9 to 10.6 percent; wages/salary income, 8.1 to 13.4 percent; and yearly property insurance, 23.9 to 25.6 percent.

The Census Bureau also compared the self-response rate differences between citizen and non-citizen households' response rates for the 2000 decennial census short form (which did not include a citizenship question) and the 2000 decennial census long form survey (the long form survey, distributed to only one in six households, included a citizenship question in 2000). Census found the decline in self-response rates for non-citizens to be 3.3 percent greater than for citizen households. However, Census was not able to isolate what percentage of decline was caused by the inclusion of a citizenship question rather than some other aspect of the long form survey (it contained over six times as many questions covering a range of topics). Indeed, the Census Bureau analysis showed that for the 2000 decennial census there was a significant drop in self response rates overall between the short and long form; the mail response rate was 66.4 percent for the short form and only 53.9 percent for the long form survey. So while there is widespread belief among many parties that adding a citizenship question could reduce response rates, the Census Bureau's analysis did not provide definitive, empirical support for that belief.

Option C, the use of administrative records rather than placing a citizenship question on the decennial census, was a potentially appealing solution to the DOJ request. The use of administrative records is increasingly part of the fabric and design of modern censuses, and the Census Bureau has been using administrative record data to improve the accuracy and reduce the cost of censuses since the early 20th century. A Census Bureau analysis matching administrative records with the 2010 decennial census and ACS responses over several more recent years showed that using administrative records could be more accurate than self-responses in the case of non-citizens. That Census Bureau analysis showed that between 28 and 34 percent of the citizenship self-responses for persons that administrative records show are non-citizens were inaccurate. In other words, when non-citizens respond to long form or ACS questions on citizenship, they inaccurately mark "citizen" about 30 percent of the time. However, the Census Bureau is still evolving its use of administrative records, and the Bureau does not yet have a complete administrative records data set for the entire population. Thus, using administrative records alone to provide DOJ with CVAP data would provide an incomplete picture. In the 2010 decennial census, the Census Bureau was able to match 88.6 percent of the population with what the Bureau considers credible administrative record data. While impressive, this means that more than 10 percent of the American population – some 25 million voting age people – would need to have their citizenship imputed by the Census Bureau. Given the scale of this number, it was imperative that another option be developed to provide a greater level of accuracy than either self-response alone or use of administrative records alone would presently provide.

I therefore asked the Census Bureau to develop a fourth alternative, **Option D**, which would combine Options B and C. Under Option D, the ACS citizenship question would be asked on the decennial census, and the Census Bureau would use the two years remaining until the 2020 decennial census to further enhance its administrative record data sets, protocols, and statistical models to provide more complete and accurate data. This approach would maximize the Census Bureau's ability to match the decennial census responses with administrative records. Accordingly, at my direction the Census Bureau is working to obtain as many additional Federal and state administrative records as possible to provide more comprehensive information for the population. "

It is my judgment that Option D will provide DOJ with the most complete and accurate CVAP data in response to its request. Asking the citizenship question of 100 percent of the population gives each respondent the opportunity to provide an answer. This may eliminate the need for the Census Bureau to have to impute an answer for millions of people. For the approximately 90 percent of the population who are citizens, this question is no additional imposition. And for the approximately 70 percent of non-citizens who already answer this question accurately on the ACS, the question is no additional imposition since census responses by law may only be used anonymously and for statistical purposes. Finally, placing the question on the decennial census and directing the Census Bureau to determine the best means to compare the decennial census responses with administrative records will permit the Census Bureau to determine the inaccurate response rate for citizens and non-citizens alike using the entire population. This will enable the Census Bureau to establish, to the best of its ability, the accurate ratio of citizen to non-citizen responses to impute for that small percentage of cases where it is necessary to do so.

Consideration of Impacts I have carefully considered the argument that the reinstatement of the citizenship question on the decennial census would depress response rate. Because a lower response rate would lead to increased non-response follow-up costs and less accurate responses, this factor was an important consideration in the decision-making process. I find that the need for accurate citizenship data and the limited burden that the reinstatement of the citizenship question would impose outweigh fears about a potentially lower response rate.

Importantly, the Department's review found that limited empirical evidence exists about whether adding a citizenship question would decrease response rates materially. Concerns about decreased response rates generally fell into the following two categories – distrust of government and increased burden. First, stakeholders, particularly those who represented immigrant constituencies, noted that members of their respective communities generally distrusted the government and especially distrusted efforts by government agencies to obtain information about them. Stakeholders from California referenced the difficulty that government agencies faced obtaining any information from immigrants as part of the relief efforts after the California wildfires. These government agencies were not seeking to ascertain the citizenship status of these wildfire victims. Other stakeholders referenced the political climate generally and fears that Census responses could be used for law enforcement purposes. But no one provided evidence that reinstating a citizenship question on the decennial census would materially decrease response rates among those who generally distrusted government and government information collection efforts, disliked the current administration, or feared law enforcement. Rather, stakeholders merely identified residents who made the decision not to participate regardless of whether the Census includes a citizenship question. The reinstatement of a citizenship question will not decrease the response rate of residents who already decided not to respond. And no one provided evidence that there are residents who would respond accurately to a decennial census that did not contain a citizenship question but would not respond if it did (although many believed that such residents had to exist). While it is possible this belief is true, there is no information available to determine the number of people who would in fact not respond due to a citizenship question being added, and no one has identified any mechanism for making such a determination.

A second concern that stakeholders advanced is that recipients are generally less likely to respond to a survey that contained more questions than one that contained fewer. The former Deputy Director and Chief Operating Officer of the Census Bureau during the George W. Bush administration described the decennial census as particularly fragile and stated that any effort to add questions risked lowering the response rate, especially a question about citizenship in the current political environment. However, there is limited empirical evidence to support this view. A former Census Bureau Director during the Obama Administration who oversaw the last decennial census noted as much. He stated that, even though he believed that the reinstatement of a citizenship question would decrease response rate, there is limited evidence to support this conclusion. This same former director noted that, in the years preceding the decennial census, certain interest groups consistently attack the census and discourage participation. While the reinstatement of a citizenship question may be a data point on which these interest groups seize in 2019, past experience demonstrates that it is likely efforts to undermine the decennial census will occur again regardless of whether the decennial census includes a citizenship question. There is no evidence that residents who are persuaded by these disruptive efforts are more or less likely to make their respective decisions about participation based specifically on the reinstatement of a citizenship question. And there are actions that the Census Bureau and stakeholder groups are taking to mitigate the impact of these attacks on the decennial census.

Additional empirical evidence about the impact of sensitive questions on survey response rates came from the SVP of Data Science at Nielsen. When Nielsen added questions on place of birth and time of arrival in the United States (both of which were taken from the ACS) to a short survey, the response rate was not materially different than it had been before these two questions were added. Similarly, the former Deputy Director and COO of the Census during the George W. Bush Administration shared an example of a citizenship-like question that he believed would negatively impact response rates but did not. He cited to the Department of Homeland Security's 2004 request to the Census Bureau to provide aggregate data on the number of Arab Americans by zip code in certain areas of the country. The Census Bureau complied, and Census employees, including the then-Deputy Director, believed that the resulting political firestorm would depress response rates for further Census Bureau surveys in the impacted communities. But the response rate did not change materially.

Two other themes emerged from stakeholder calls that merit discussion. First, several stakeholders who opposed reinstatement of the citizenship question did not appreciate that the question had been asked in some form or another for nearly 200 years. Second, other stakeholders who opposed reinstatement did so based on the assumption that the data on citizenship that the Census Bureau collects through the ACS are accurate, thereby obviating the need to ask the question on the decennial census. But as discussed above, the Census Bureau estimates that between 28 and 34 percent of citizenship self-responses on the ACS for persons that administrative records show are non-citizens were inaccurate. Because these stakeholder concerns were based on incorrect premises, they are not sufficient to change my decision.

Finally, I have considered whether reinstating the citizenship question on the 2020 Census will lead to any significant monetary costs, programmatic or otherwise. The Census Bureau staff have advised that the costs of preparing and adding the question would be minimal due in large part to the fact that the citizenship question is already included on the ACS, and thus the citizenship question has already undergone the cognitive research and questionnaire testing required for new questions. Additionally, changes to the Internet Self-Response instrument, revising the Census Questionnaire Assistance, and redesigning of the printed questionnaire can be easily implemented for questions that are finalized prior to the submission of the list of questions to Congress.

The Census Bureau also considered whether non-response follow-up increases resulting from inclusion of the citizenship question would lead to increased costs. As noted above, this estimate was difficult to assess given the Census Bureau and Department's inability to determine what impact there will be on decennial census survey responses. The Bureau provided a rough estimate that postulated that up to 630,000 additional households may require NRFU operations if a citizenship question is added to the 2020 decennial census. However, even assuming that estimate is correct, this additional ½ percent increase in NRFU operations falls well within the margin of error that the Department, with the support of the Census Bureau, provided to Congress in the revised Lifecycle Cost Estimate ("LCE") this past fall. That LCE assumed that NRFU operations might increase by 3 percent due to numerous factors, including a greater increase in citizen mistrust of government, difficulties in accessing the Internet to respond, and other factors.

Inclusion of a citizenship question on this country's decennial census is not new – the decision to collect citizenship information from Americans through the decennial census was first made centuries ago. The decision to include a citizenship question on a national census is also not uncommon. The United Nations recommends that its member countries ask census questions identifying both an individual's country of birth and the country of citizenship. *Principals and Recommendations for Population and Housing Censuses (Revision 3)*, UNITED NATIONS 121 (2017). Additionally, for countries in which the population may include a large portion of naturalized citizens, the United Nations notes that, "it may be important to collect information on the method of acquisition of citizenship." *Id.* at 123. And it is important to note that other major democracies inquire about citizenship on their census, including Australia, Canada, France, Germany, Indonesia, Ireland, Mexico, Spain, and the United Kingdom, to name a few.

The Department of Commerce is not able to determine definitively how inclusion of a citizenship question on the decennial census will impact responsiveness. However, even if there is some impact on responses, the value of more complete and accurate data derived from surveying the entire population outweighs such concerns. Completing and returning decennial census questionnaires is required by Federal law, those responses are protected by law, and inclusion of a citizenship question on the 2020 decennial census will provide more complete information for those who respond. The citizenship data provided to DOJ will be more accurate with the question than without it, which is of greater importance than any adverse effect that may result from people violating their legal duty to respond.

To conclude, after a thorough review of the legal, program, and policy considerations, as well as numerous discussions with the Census Bureau leadership and interested stakeholders, I have determined that reinstatement of a citizenship question on the 2020 decennial census is necessary to provide complete and accurate data in response to the *DOJ* request. To minimize any impact on decennial census response rates, I am directing the Census Bureau to place the citizenship question last on the decennial census form.

Please make my decision known to Census Bureau personnel and Members of Congress prior to March 31, 2018. I look forward to continuing to work with the Census Bureau as we strive for a complete and accurate 2020 decennial census.

CC: Ron Jarmin, performing the nonexclusive functions and duties of the Director of the Census Bureau

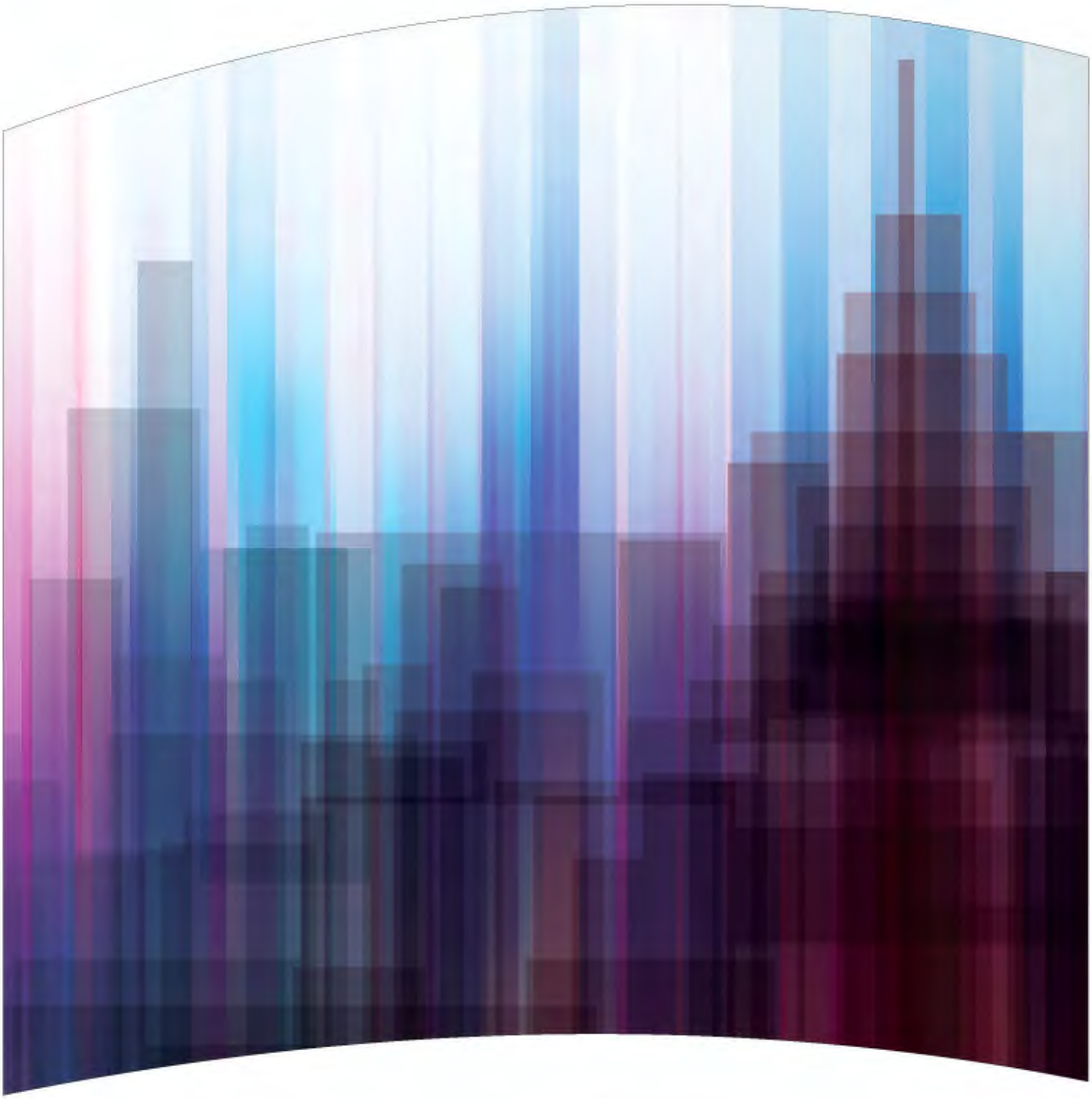
Enrique Lamas, performing the nonexclusive functions and duties of the Deputy Director of the Census Bureau

Exhibit 2

Questions Planned for the 2020 Census and American Community Survey

Federal Legislative and Program Uses

Issued March 2018



United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

This page is intentionally blank.

Contents

Introduction . . . 1

Protecting the Information Collected by These Questions . . . 2

Questions Planned for the 2020 Census . . . 3

- Age . . . 5
- Citizenship . . . 7
- Hispanic Origin . . . 9
- Race . . . 11
- Relationship . . . 13
- Sex . . . 15
- Tenure (Owner/Renter) . . . 17
- Operational Questions for use in the 2020 Census . . . 19

Questions Planned for the American Community Survey . . . 21

- Acreage and Agricultural Sales . . . 23
- Age . . . 25
- Ancestry . . . 27
- Commuting (Journey to Work) . . . 29
- Computer and Internet Use . . . 31
- Disability . . . 33
- Fertility . . . 35
- Grandparent Caregivers . . . 37
- Health Insurance Coverage and Health Insurance Premiums and Subsidies . . . 39
- Hispanic Origin . . . 41
- Home Heating Fuel . . . 43
- Home Value and Rent . . . 45
- Income . . . 47
- Industry, Occupation, and Class of Worker . . . 49
- Labor Force Status . . . 51
- Language Spoken at Home . . . 53
- Marital Status and Marital History . . . 55
- Migration (Previous Residence)/Residence 1 Year Ago . . . 57
- Place of Birth, Citizenship, and Year of Entry . . . 59
- Plumbing Facilities, Kitchen Facilities, and Telephone Service . . . 61
- Race . . . 63
- Relationship . . . 65
- School Enrollment, Educational Attainment, and Undergraduate Field of Degree . . . 67
- Selected Monthly Owner Costs . . . 69
- Sex . . . 71
- Supplemental Nutrition Assistance Program (SNAP)/Food Stamps . . . 73
- Tenure (Owner/Renter) . . . 75
- Units in Structure, Rooms, and Bedrooms . . . 77
- Vehicles Available . . . 79
- Veteran Status, Period of Military Service, and Department of Veterans Affairs (VA) Service-Connected Disability Rating . . . 81
- Work Status Last Year . . . 83
- Year Built and Year Moved In . . . 85
- Operational Questions . . . 87

Appendix A: Year the Planned Questions Were First Asked in the Decennial Census Program . . . A1

This page is intentionally blank.

Introduction

BACKGROUND

Since 1790, a national census of the U.S. population has been conducted every 10 years, as required by the U.S. Constitution. Additional information beyond the population count has been collected with each census in response to the challenges facing the nation and a national desire to understand ourselves.

For much of the 20th century, most addresses received a "short" form, while a portion of addresses received a more detailed "long" form. The Census 2000 short form was designed to collect basic demographic and housing information (i.e., age, sex, race, Hispanic origin, relationship, and tenure) to be used for congressional apportionment and to provide data to states for intrastate redistricting. The long form, sent to approximately 1 in 6 households, collected social, housing, and economic information (e.g., educational attainment, disability status, employment status, income, and housing costs) that was used to plan and determine funding for a wide array of federal, state, local, and tribal programs.

Since 2005, in order to provide communities, businesses, and the public with the detailed long-form information more frequently, these data have been collected monthly (and released annually) through the American Community Survey (ACS).[1] This innovation enabled the 2010 Census to be a "short-form-only" census. Decoupling the collection of short- and long-form data allowed the U.S. Census Bureau to focus decennial census efforts on the constitutional requirements to produce a count of the resident population, while employing technology in both collections to improve efficiencies, improve accuracy, and reduce costs. The result has been the dissemination of more current and detailed information than has ever been available.

The 2020 Decennial Census Program, comprised of the 2020 Census and the ACS, will provide an official count through a "short-form-only" census (i.e., age, sex, race, Hispanic origin, relationship, tenure, and citizenship), as well as a portrait of communities counted across the nation through data collected by the ACS. This program is the only data-gathering effort that collects information from enough people to produce comparable data for every geographic area recognized by the Census Bureau.

SUBMISSION OF QUESTIONS PLANNED FOR THE 2020 DECENNIAL CENSUS PROGRAM

The Census Act requires that the questions to be included in the next census be submitted to Congress no later than 2 years before the census date. The contents of this handbook describe the questions that will be asked on the 2020 Census and the ACS.

ABOUT THE QUESTIONS PLANNED FOR THE 2020 DECENNIAL CENSUS PROGRAM

Throughout each decade, regular content reviews are conducted to ensure that the information collected through the decennial census program is required by federal programs.[2]

To prepare for the 2020 Census, the Office of Management and Budget and the Census Bureau embarked on a comprehensive review including chartering the Interagency Council on Statistical Policy (ICSP) Subcommittee on the ACS and conducting the 2014 ACS Content Review. This effort was designed to examine and confirm the value of each question on the ACS, and to confirm and update the statutory and regulatory authority for the questions with federal agencies. In 2016, the Census Bureau asked federal agencies to provide any updates to this documentation.

The resulting information about federal uses is presented throughout the descriptions of the questions on the following pages. These descriptions are designed to give the reader a clear understanding of 1) the relationship between questions asked of respondents and the summarized data that are released in published tables, 2) how federal agencies use the resulting data, and 3) the benefits of the data at the community level.

[1] The ACS also collects short-form data on its questionnaire. However, the ACS asks for basic demographic and housing information from a sample of households, while the decennial census asks for basic demographic and housing information from all households.

[2] Selected statutory uses of each subject are published in the "Subjects Planned for the 2020 Census and American Community Survey," available at <www.census.gov/library/publications/2017/dec/planned-subjects-2020-acs.html>.

Protecting the Information Collected by These Questions

The Census Bureau has an obligation to produce accurate, relevant statistics about the nation's economy and people, but we recognize that the information collected by these questions is often private. We depend on cooperation and trust, and promise to protect the confidentiality of this information.

Federal law protects this information; Title 13 of the U.S. Code protects the confidentiality of all collected information. Violating this law is a crime with severe penalties. Please visit <www.census.gov/about/policies/privacy/data_protection/federal_law.html> for details.

OUR PRIVACY PRINCIPLES

We recognize the value of respondent trust, and we believe that when a person answers the 2020 Census or the ACS we must serve as caretakers of the information. The Census Bureau's Privacy Principles remind us of this promise and help ensure the protection of respondent information throughout all of our activities.

The Privacy Principles are our guidelines. They help us as we determine content to consider respondents' rights and concerns. Every principle embodies a promise to the respondent.

Necessity: Do we need to collect information on this question?

Every time we prepare to ask a question, we determine whether the information is truly necessary. All of the information we collect is used for federal programs.

- We promise to collect only information necessary for each survey and census.
- We promise that we will use the information only to produce timely, relevant statistics about the population and the economy of the United States.

Openness: Do respondents know why we are collecting this information?

We collect information only for statistical purposes, and it is never used to identify individuals. Before participating, respondents have the right to know why we are conducting the survey or census, why we are asking specific questions, and the purposes for which the information will be used.

- We promise to inform respondents about the purpose and uses for every survey or census we conduct before respondents provide answers.

Respectful treatment of respondents: Are our efforts reasonable and do we treat people with respect?

- We promise to minimize the effort and time it takes for respondents to participate in the data collection by efficient designs.
- We promise to use only legal, ethical, and professionally accepted practices in collecting data.
- We promise to ensure any collection of sensitive information from children and other sensitive populations does not violate federal protections for research participants and is done only when it benefits the public good.

Confidentiality: How do we protect this information?

In addition to removing personally identifiable information (i.e., names, telephone numbers, and addresses) from our data files, we use various approaches to protect personal information—including computer technologies, statistical methodologies, and security procedures.

Our security measures ensure that only a restricted number of authorized people have access to private information and that access is only granted to conduct our work and for no other purposes. Every person who works with census confidential information collected by the Census Bureau is sworn for life to uphold the law.

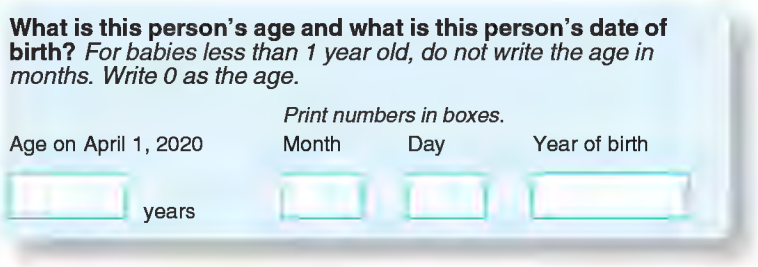
Violating the confidentiality of a respondent is a federal crime with serious penalties, including a federal prison sentence of up to 5 years, a fine of up to $250,000, or both.

- We promise that every person with access to respondent information is sworn for life to protect respondent confidentiality.
- We promise that we will use every technology, statistical methodology, and physical security procedure at our disposal to protect respondent information.

Questions Planned for the 2020 Census

This page is intentionally blank.

Age asked since 1790.



ANSWERS TO THE AGE AND DATE OF BIRTH QUESTION PROVIDE THE DATA THAT HELP US UNDERSTAND THE SIZE OF DIFFERENT AGE GROUPS AND HOW OTHER CHARACTERISTICS MAY VARY BY AGE.

Age data are used in planning and funding government programs that provide funds or services for specific age groups, such as children, working-age adults, women of childbearing age, or the older population. These statistics are also used to enforce laws, regulations, and policies against age discrimination in government programs and in society.

AGE DATA HELP COMMUNITIES:

Provide Assistance to Older Americans

Knowing how many people in a community are aged 60 and older helps local officials provide programs and services that enable older adults to remain living safely in their homes and communities (Older Americans Act). Age data are also used in programs that provide services and assistance to seniors, such as financial assistance with utilities (Low Income Home Energy Assistance Program).

Provide Assistance to Children and Families

Knowing the numbers and ages of children in families can help communities enroll eligible families in programs designed to assist them. For example, age data are used in targeted efforts to enroll eligible people in Medicaid and the Children's Health Insurance Program.

Educate Children and Adults

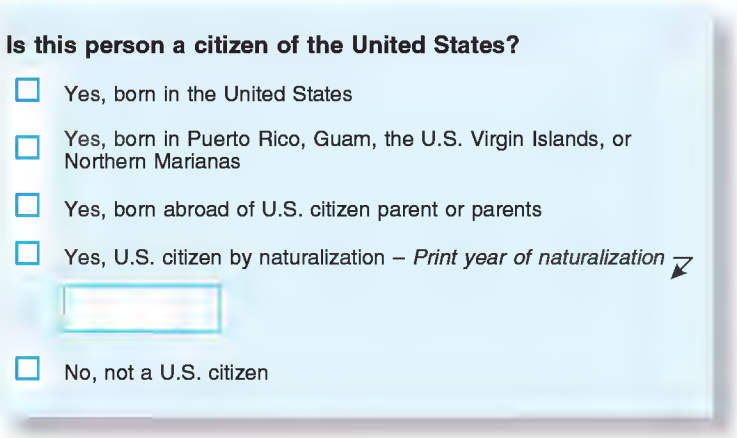
Knowing how many children and adults depend on services through schools helps school districts make long-term building, staffing, and funding decisions.

Ensure Equal Opportunity

Knowing the age distribution of a community's population helps government and communities enforce laws, regulations, and policies against discrimination based on age.

This page is intentionally blank.

Citizenship asked since 1820.[1]



A QUESTION ABOUT A PERSON'S CITIZENSHIP IS USED TO CREATE STATISTICS ABOUT CITIZEN AND NONCITIZEN POPULATIONS.

These statistics are essential for enforcing the Voting Rights Act and its protections against voting discrimination.

CITIZENSHIP DATA HELP COMMUNITIES:

Enforce Voting Rights Law

Knowing how many people reside in the community and how many of those people are citizens, in combination with other information, provides the statistical information that helps the government enforce Section 2 of the Voting Rights Act and its protections against discrimination in voting.

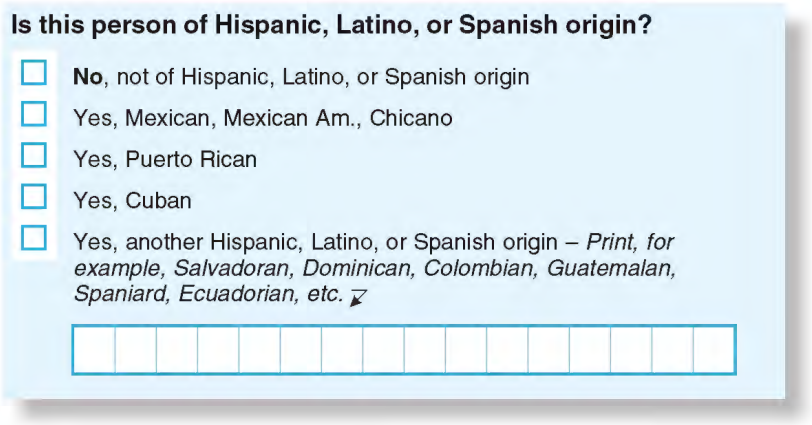
Understand Changes

Knowing how many citizens and noncitizens live in the United States, in combination with other information, is of interest to researchers, advocacy groups, and policymakers.

[1] Citizenship asked 1820, 1830, 1870, and 1890 to present.

This page is intentionally blank.

Hispanic Origin asked since 1970.



A QUESTION ABOUT WHETHER A PERSON IS OF HISPANIC, LATINO, OR SPANISH ORIGIN IS USED TO CREATE STATISTICS ABOUT THIS ETHNIC GROUP.

These data are required for federal and state programs and are critical factors in the basic research behind numerous policies, particularly for civil rights. Hispanic origin data are used in planning and funding government programs that provide funds or services for specific groups. These data are also used to evaluate government programs and policies to ensure they fairly and equitably serve the needs of the Hispanic population and to monitor compliance with antidiscrimination laws, regulations, and policies. States also use these data to meet legislative redistricting requirements.

Though many respondents expect to see a Hispanic, Latino, or Spanish category on the race question, this question is asked separately because people of Hispanic origin may be of any race(s) in accordance with the 1997 Office of Management and Budget standards on race and ethnicity.

HISPANIC ORIGIN DATA HELP COMMUNITIES:

Ensure Equal Opportunity

We ask about the Hispanic or Latino origin of community members to help governments and communities enforce antidiscrimination laws, regulations, and policies. For example, data on the Hispanic population are used to:

- Establish and evaluate the guidelines for federal affirmative action plans under the Federal Equal Opportunity Recruitment Program.
- Monitor compliance with the Voting Rights Act.
- Monitor and enforce equal employment opportunities under the Civil Rights Act of 1964.
- Identify segments of the population who may not be getting needed medical services under the Public Health Service Act.

Understand Changes

Researchers, advocacy groups, and policymakers are interested in knowing if the distribution of the Hispanic and non-Hispanic population changes by age, sex, relationship, and housing tenure.

This page is intentionally blank.

Race asked since 1790.

What is this person's race?
Mark [X] one or more boxes ***AND*** *print origins.*

- ☐ White – *Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc.*
- ☐ Black or African Am. – *Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.*
- ☐ American Indian or Alaska Native – *Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.*
- ☐ Chinese
- ☐ Filipino
- ☐ Asian Indian
- ☐ Vietnamese
- ☐ Korean
- ☐ Japanese
- ☐ Other Asian – *Print, for example, Pakistani, Cambodian, Hmong, etc.*
- ☐ Native Hawaiian
- ☐ Samoan
- ☐ Chamorro
- ☐ Other Pacific Islander – *Print, for example, Tongan, Fijian, Marshallese, etc.*
- ☐ Some other race – *Print race or origin.*

A QUESTION ABOUT A PERSON'S RACE IS USED TO CREATE STATISTICS ABOUT RACE AND TO PRESENT OTHER STATISTICS BY RACE GROUPS.

These data are required for federal, state, and tribal programs and are critical factors in the basic research behind numerous policies, particularly for civil rights. Race data are used in planning and funding government programs that provide funds or services for specific groups. These data are also used to evaluate government programs and policies to ensure they fairly and equitably serve the needs of all racial groups and to monitor compliance with antidiscrimination laws, regulations, and policies. States also use these data to meet legislative redistricting requirements.

The U.S. Census Bureau collects race data in accordance with the 1997 Office of Management and Budget standards on race and ethnicity. The categories on race are based on self-identification and generally reflect a social definition of race. The categories are not an attempt to define race biologically, anthropologically, or genetically.

Race asked since 1790—Con.

RACE DATA HELP COMMUNITIES:

Ensure Equal Opportunity

Knowing the races of community members helps government and communities enforce antidiscrimination laws, regulations, and policies. For example, race data are used in the following ways:

- Establish and evaluate the guidelines for federal affirmative action plans under the Federal Equal Opportunity Recruitment Program.
- Monitor compliance with the Voting Rights Act.
- Monitor and enforce equal employment opportunities under the Civil Rights Act of 1964.
- Identify segments of the population who may not be getting needed medical services under the Public Health Service Act.

Understand Changes

Researchers, advocacy groups, and policymakers are interested in knowing if the distribution of the different racial groups changes by age, sex, relationship, and housing tenure.

Administer Programs for Specific Groups

Knowing how many people are eligible to participate in certain programs helps communities, including tribal governments, ensure that programs are operating as intended. For example, the Indian Housing Block Grant program, Indian Community Development Block Grant program, and Indian Health Service all depend on accurate statistics of American Indians and Alaska Natives. Data for the American Indian and Alaska Native population come from the question about a person's race.

Relationship asked since 1880.

How is this person related to Person 1? *Mark ☒ ONE box.*

- ☐ Opposite-sex husband/wife/spouse
- ☐ Opposite-sex unmarried partner
- ☐ Same-sex husband/wife/spouse
- ☐ Same-sex unmarried partner
- ☐ Biological son or daughter
- ☐ Adopted son or daughter
- ☐ Stepson or stepdaughter
- ☐ Brother or sister
- ☐ Father or mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law or daughter-in-law
- ☐ Other relative
- ☐ Roommate or housemate
- ☐ Foster child
- ☐ Other nonrelative

A QUESTION ABOUT THE RELATIONSHIP OF EACH PERSON IN A HOUSEHOLD TO ONE CENTRAL PERSON IS USED TO CREATE ESTIMATES ABOUT FAMILIES, HOUSEHOLDS, AND OTHER GROUPS, AND TO PRESENT OTHER DATA AT A HOUSEHOLD LEVEL.

Relationship data are used in planning and funding government programs that provide funds or services for families, people living or raising children alone, grandparents living with grandchildren, or other households that qualify for additional assistance.

RELATIONSHIP DATA HELP COMMUNITIES:

Provide Adequate Housing

Knowing about the different types of households in a community (single people, couples, families, roommates, etc.) helps communities understand whether available housing meets the needs of residents.

When housing is not sufficient or not affordable, relationship data can help communities enroll eligible households in programs designed to assist them, and can help communities qualify for grants from the Community Development Block Grant, HOME Investment Partnership Program, Emergency Solutions Grant, Housing Opportunities for Persons With AIDS, and other programs.

Provide Assistance to Families

Knowing more about families, such as the ages of children, can help communities enroll eligible families in programs designed to assist them, such as Head Start and the Children's Health Insurance Program, and can help communities qualify for grants to fund these programs. Relationship data are also used to ensure that programs like Temporary Assistance for Needy Families are making a difference for families.

Relationship asked since 1880—Con.

Understand Changing Households

Information about living arrangements and how they are changing, including whether older residents are staying in their homes as they age, whether young people are living with parents or moving in with roommates, and which kinds of households include young children, can help communities plan future programs and services for residents. For example, the Social Security Administration estimates future program needs based on the current relationships of working people.

Sex asked since 1790.

What is this person's sex? *Mark [X] ONE box.*

☐ Male ☐ Female

A QUESTION ABOUT THE SEX OF EACH PERSON IS USED TO CREATE STATISTICS ABOUT MALES AND FEMALES AND TO PRESENT OTHER DATA BY SEX.

Data disaggregated by sex are used in planning and funding government programs and in evaluating other government programs and policies to ensure they fairly and equitably serve the needs of males and females. These statistics are also used to enforce laws, regulations, and policies against discrimination in government programs and in society.

DATA DISAGGREGATED BY SEX HELP COMMUNITIES:

Provide Resources for Education

Data disaggregated by sex are used to allocate funds to institutions of higher learning that increase participation, particularly of minority women, in scientific and engineering programs under the Higher Education Act.

Ensure Equal Opportunity

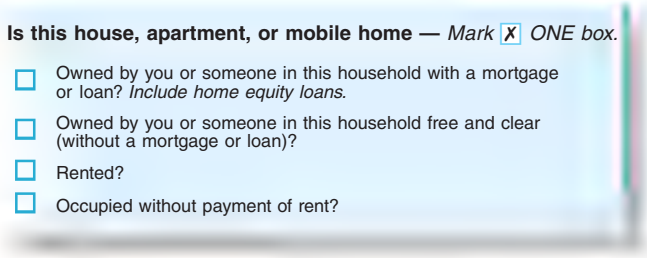
Data disaggregated by sex are used to evaluate employment practices under the Civil Rights Act of 1964. The Equal Employment Opportunity Commission, using sex-disaggregated data, establishes and evaluates guidelines dealing with equal employment opportunity. Data disaggregated by sex are used to evaluate housing policies and practices under the Civil Rights Act.

Understand Changes

State and local planners analyze social and economic characteristics of males and females to predict future needs for housing, education, childcare, health care, transportation, and employment.

This page is intentionally blank.

Tenure (Owner/Renter) asked since 1890.



Is this house, apartment, or mobile home — *Mark ☒ ONE box.*

- ☐ Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*
- ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
- ☐ Rented?
- ☐ Occupied without payment of rent?

A QUESTION ABOUT WHETHER A HOME IS OWNED OR RENTED IS USED TO CREATE DATA ABOUT TENURE, RENTERS, AND HOMEOWNERSHIP.

Tenure is the most basic characteristic to assess housing inventory. Tenure data are used in government programs that analyze whether adequate housing is available to residents. Tenure data are also used to provide and fund housing assistance programs. These statistics are also used to enforce laws, regulations, and policies against discrimination in private-market housing, government programs, and in society.

TENURE DATA HELP COMMUNITIES:

Provide Adequate Housing

Knowing the rates of home rental and ownership helps communities understand whether available housing meets the needs of residents.

When housing is not sufficient, data about owners and renters can help communities enroll eligible households in programs designed to assist them.

Plan Community Development

Knowing how the balance of rented homes, mortgaged homes, and homes owned free and clear changes over time can help communities understand changes in local housing markets; identify opportunities to improve tax, assistance, and zoning policies; and reduce tax revenue losses from vacant or abandoned properties.

Ensure Equal Opportunity

Knowing the characteristics of people who rent and people who own homes in the community, such as age, sex, race, and Hispanic origin, helps government and communities enforce laws, such as the 1968 Fair Housing Act, designed to eliminate discrimination in housing.

Understand Changing Households

Knowing whether older residents are staying in homes as they age or moving into rented homes, and whether young people are staying with parents, renting with roommates, or buying homes, can help governments and communities distribute funds appropriately between homeownership and rental housing programs and services for residents.

This page is intentionally blank.

Operational Questions for use in the 2020 Census.

How many people were living or staying in this house, apartment, or mobile home on April 1, 2020?

Number of people = ☐☐

Were there any additional people staying here on April 1, 2020 that you did not include in Question 1?

Mark ☒ all that apply.

- ☐ Children, related or unrelated, such as newborn babies, grandchildren, or foster children
- ☐ Relatives, such as adult children, cousins, or in-laws
- ☐ Nonrelatives, such as roommates or live-in babysitters
- ☐ People staying here temporarily
- ☐ No additional people

What is your telephone number?

We will only contact you if needed for official Census Bureau business.

Telephone Number

☐☐☐ – ☐☐☐ – ☐☐☐☐

Please provide information for each person living here. If there is someone living here who pays the rent or owns this residence, start by listing him or her as Person 1. If the owner or the person who pays the rent does not live here, start by listing any adult living here as Person 1.

What is Person 1's name? *Print name below.*

First Name MI

Last Name(s)

Does this person usually live or stay somewhere else?

Mark ☒ all that apply.

- ☐ No
- ☐ Yes, for college
- ☐ Yes, for a military assignment
- ☐ Yes, for a job or business
- ☐ Yes, in a nursing home
- ☐ Yes, with a parent or other relative
- ☐ Yes, at a seasonal or second residence
- ☐ Yes, in a jail or prison
- ☐ Yes, for another reason

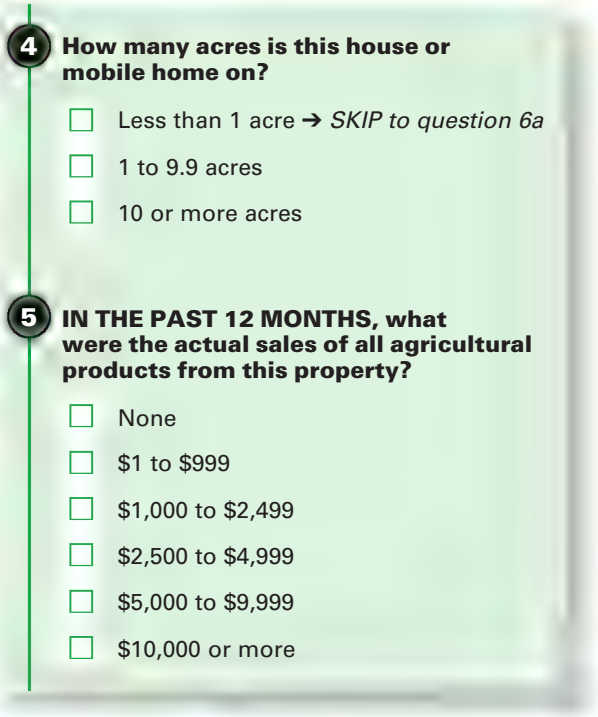
Some operational questions will appear on the 2020 Census that will not result in published counts. These questions are asked to better administer the data collection process and to ensure greater accuracy of the data collected. Contact information is not part of published estimates and is carefully protected, as mandated by federal law, to respect the personal information of respondents.

This page is intentionally blank.

Questions Planned for the American Community Survey

This page is intentionally blank.

Acreage asked since 1960.
Agricultural Sales asked since 1960.



4 **How many acres is this house or mobile home on?**

- ☐ Less than 1 acre ➔ *SKIP to question 6a*
- ☐ 1 to 9.9 acres
- ☐ 10 or more acres

5 **IN THE PAST 12 MONTHS, what were the actual sales of all agricultural products from this property?**

- ☐ None
- ☐ $1 to $999
- ☐ $1,000 to $2,499
- ☐ $2,500 to $4,999
- ☐ $5,000 to $9,999
- ☐ $10,000 or more

QUESTIONS ABOUT THE ACREAGE ASSOCIATED WITH HOUSES, MOBILE HOMES, AND AGRICULTURAL SALES ARE USED TO CREATE DATA ABOUT AGRICULTURAL PROPERTIES AND TO BETTER UNDERSTAND HOME VALUE STATISTICS.

These data are used in planning government programs designed to benefit the farm population and identifying or excluding agricultural areas for many other programs.

ACREAGE AND AGRICULTURAL SALES DATA HELP COMMUNITIES:

Provide Equitable Housing Assistance

Knowing which homes might qualify for farm subsidies and which homes qualify for housing subsidies is important to ensure that funds are fairly allocated. For example, the historical definition of Fair Market Rents, used to allocate housing assistance, has always excluded units on acreage of more than 10 acres to eliminate those units that might benefit from farm subsidies and therefore have lower-than-market rents. Understanding which kinds of properties are eligible for certain programs helps communities inform eligible residents and determine whether the community is eligible for funds based on its farm population.

Support Agricultural Programs

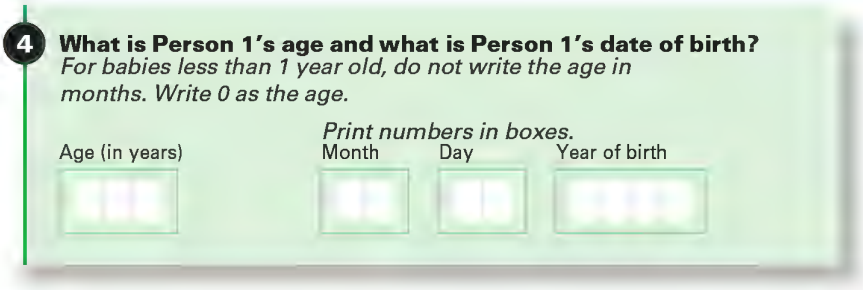
Knowing which areas of a community are agricultural helps communities ensure eligible institutions receive funding for cooperative agricultural extension work and agricultural research. This funding is distributed to eligible institutions based on a legislatively determined formula that uses these data.

Plan Community Development

Knowing the size and agricultural nature of areas of each community can help communities understand changes in local housing markets; identify opportunities to improve tax, assistance, and zoning policies; and reduce tax revenue losses from vacant or abandoned properties.

This page is intentionally blank.

Age asked since 1790.



AGE AND DATE OF BIRTH QUESTIONS ARE USED TO UNDERSTAND THE SIZE AND CHARACTERISTICS OF DIFFERENT AGE GROUPS AND TO PRESENT OTHER DATA BY AGE.

Age data are used in planning and funding government programs that provide funds or services for specific age groups, such as children, working-age adults, women of childbearing age, or the older population. These statistics are also used to enforce laws, regulations, and policies against age discrimination in government programs and in society.

AGE DATA HELP COMMUNITIES:

Provide Assistance to Older Americans

Knowing how many people in a community are aged 60 and older helps local officials provide programs and services that enable older adults to remain living safely in their homes and communities (Older Americans Act). Age data are also used in programs that provide services and assistance to seniors, such as financial assistance with utilities (Low Income Home Energy Assistance Program).

Provide Assistance to Children and Families

Knowing the numbers and ages of children in families in combination with other information, such as household income, health insurance status, and poverty status, can help communities enroll eligible families in programs designed to assist them. For example, age data are used in targeted efforts to enroll eligible people in Medicaid and the Children's Health Insurance Program.

Educate Children and Adults

Knowing how many children and adults depend on services through schools helps school districts make long-term building, staffing, and funding decisions. Age in combination with other information, such as disability status, language spoken at home, and poverty status, assists schools in understanding the needs of their students and qualifying for grants that help fund programs for those students (Elementary and Secondary Education Act of 1965).

Ensure Equal Opportunity

Knowing the ages of people in the community in combination with information about housing, employment, and education, helps government and communities enforce laws, regulations, and policies against discrimination based on age. For example, age information is used to analyze the employment status of workers by age (Age Discrimination in Employment Act).

This page is intentionally blank.

Ancestry asked since 1980.

13 **What is this person's ancestry or ethnic origin?**

(For example: Italian, Jamaican, African Am., Cambodian, Cape Verdean, Norwegian, Dominican, French Canadian, Haitian, Korean, Lebanese, Polish, Nigerian, Mexican, Taiwanese, Ukrainian, and so on.)

A QUESTION ABOUT A PERSON'S ANCESTRY OR ETHNIC ORIGIN IS USED TO CREATE STATISTICS ABOUT VARIOUS ANCESTRY GROUPS.

Ancestry data are used in planning and evaluating government programs and policies to ensure they fairly and equitably serve the needs of all groups. These statistics are also used to enforce laws, regulations, and policies against discrimination in society.

ANCESTRY DATA HELP COMMUNITIES:

Ensure Equal Opportunity

Knowing the ethnic groups in a community in combination with information about housing, voting, language, employment, and education, helps government and communities enforce laws, regulations, and policies against discrimination based on national origin. For example, ancestry data are used to enforce nondiscrimination in education (including monitoring desegregation); to enforce nondiscrimination in employment by federal agencies, private employers, employment agencies, and labor organizations; and to enforce laws, regulations, and policies against discrimination in federal financial assistance (Civil Rights Act of 1964).

Understand Changes

Knowing whether people from different backgrounds have the same opportunities in education, employment, voting, homeownership, and many other areas is of interest to researchers, advocacy groups, and policymakers. For example, ancestry data are used with age and language data to address language and cultural diversity needs in health care plans for the older population.

This page is intentionally blank.

Commuting (Journey to Work) asked since 1960.

31 **At what location did this person work LAST WEEK?** *If this person worked at more than one location, print where he or she worked most last week.*

a. Address (Number and street name)

If the exact address is not known, give a description of the location such as the building name or the nearest street or intersection.

b. Name of city, town, or post office

c. Is the work location inside the limits of that city or town?

- ☐ Yes
- ☐ No, outside the city/town limits

d. Name of county

e. Name of U.S. state or foreign country

f. ZIP Code

32 **How did this person usually get to work LAST WEEK?** *Mark ONE box for the method of transportation used for most of the distance.*

- ☐ Car, truck, or van
- ☐ Bus
- ☐ Subway or elevated rail
- ☐ Long-distance train or commuter rail
- ☐ Light rail, streetcar, or trolley
- ☐ Ferryboat
- ☐ Taxicab
- ☐ Motorcycle
- ☐ Bicycle
- ☐ Walked
- ☐ Worked from home → *SKIP to question 40a*
- ☐ Other method

33 **How many people, including this person, usually rode to work in the car, truck, or van LAST WEEK?**

Person(s)

34 **LAST WEEK, what time did this person's trip to work usually begin?**

Hour : Minute

- ☐ a.m.
- ☐ p.m.

35 **How many minutes did it usually take this person to get from home to work LAST WEEK?**

Minutes

QUESTIONS ABOUT WHERE PEOPLE WORK, HOW THEY GET THERE, WHEN THEY LEAVE, AND HOW LONG IT TAKES ARE USED TO CREATE DATA ABOUT COMMUTING OR A PERSON'S JOURNEY TO WORK.

Journey to work data are used in planning and funding for improvements to road and highway infrastructure, developing transportation plans and services, and understanding where people are traveling in the course of a normal day. These data are also used to evaluate transportation plans to ensure they fairly and equitably serve the needs of all groups.

Commuting (Journey to Work) asked since 1960—Con.

COMMUTING DATA HELP COMMUNITIES:

Improve Transportation Planning

Knowing where people commute to and from, and what time of day they are commuting, helps transportation planners create mass transportation and metropolitan transportation plans that are compliant with various transportation, environmental, and antidiscrimination regulations.

Local agencies and organizations use these statistics to plan transportation programs and services that meet the diverse needs of local populations, including the disabled population, bicycle commuters, carpools, and ride-sharers. Commuting data are also used to forecast future use of new or updated transportation systems.

Ensure Equal Opportunity

Knowing where people could reasonably commute from in order to work in a certain area is used by communities and businesses for employment planning, and by communities and governments to enforce laws, regulations, and policies against employment discrimination.

Understand Changes in Commutes

As commuting patterns change, information about where people could reasonably commute from in order to work in a certain area is used to understand commercial markets and labor force participation, and to plan local emergency response programs.

Computer and Internet Use asked since 2013.

9 **At this house, apartment, or mobile home – do you or any member of this household own or use any of the following types of computers?**

	Yes	No
a. Desktop or laptop	☐	☐
b. Smartphone	☐	☐
c. Tablet or other portable wireless computer	☐	☐
d. Some other type of computer *Specify*	☐	☐

10 **At this house, apartment, or mobile home – do you or any member of this household have access to the internet?**

- ☐ Yes, by paying a cell phone company or internet service provider
- ☐ Yes, without paying a cell phone company or internet service provider → *SKIP to question 12*
- ☐ No access to the internet at this house, apartment, or mobile home → *SKIP to question 12*

11 **Do you or any member of this household have access to the internet using a –**

	Yes	No
a. cellular data plan for a smartphone or other mobile device?	☐	☐
b. broadband (high speed) internet service such as cable, fiber optic, or DSL service installed in this household?	☐	☐
c. satellite internet service installed in this household?	☐	☐
d. dial-up internet service installed in this household?	☐	☐
e. some other service? *Specify service*	☐	☐

QUESTIONS ABOUT THE COMPUTERS AND DEVICES THAT PEOPLE USE, WHETHER PEOPLE ACCESS THE INTERNET, AND HOW PEOPLE ACCESS THE INTERNET ARE USED TO CREATE DATA ABOUT COMPUTER AND INTERNET USE.

These statistics were first released to the public in September 2014. The questions were added as a requirement of the Broadband Data Improvement Act of 2008. They help federal agencies measure the nationwide development of broadband access and decrease barriers to broadband access.

COMPUTER AND INTERNET USE DATA HELP COMMUNITIES:

Ensure Residents Can Communicate

State and local agencies can use these statistics to evaluate access to broadband in their communities. Schools, libraries, rural health care providers, and other public service providers also use this information. Communities ensure their residents are connected to assistance programs, emergency services, and important information. These statistics may also be useful to understand whether to use Internet or more expensive outreach methods for distributing important public health or safety information.

Federal agencies use these data to evaluate the extent of access to, and adoption of broadband, with a focus on underserved areas. State and local agencies might choose to use these statistics to evaluate access to broadband in their communities.

This page is intentionally blank.

Disability asked since 1830.

18 **a. Is this person deaf or does he/she have serious difficulty hearing?**

- ☐ Yes
- ☐ No

b. Is this person blind or does he/she have serious difficulty seeing even when wearing glasses?

- ☐ Yes
- ☐ No

19 **a. Because of a physical, mental, or emotional condition, does this person have serious difficulty concentrating, remembering, or making decisions?**

- ☐ Yes
- ☐ No

b. Does this person have serious difficulty walking or climbing stairs?

- ☐ Yes
- ☐ No

c. Does this person have difficulty dressing or bathing?

- ☐ Yes
- ☐ No

20 **Because of a physical, mental, or emotional condition, does this person have difficulty doing errands alone such as visiting a doctor's office or shopping?**

- ☐ Yes
- ☐ No

QUESTIONS ABOUT A PERSON'S DIFFICULTY WITH SPECIFIC DAILY TASKS ARE USED TO CREATE STATISTICS ABOUT DISABILITY.

Disability data are used in planning and funding government programs that provide funds or services for populations with disabilities. In addition, these data are used in evaluating other government programs and policies to ensure that they fairly and equitably serve the needs of all groups. These statistics are also used to enforce laws, regulations, and policies against discrimination.

DISABILITY DATA HELP COMMUNITIES:

Provide Adequate Housing

Knowing the different types of disabled households in a community helps communities understand whether available housing meets the needs of residents. When housing is not sufficient or not affordable, disability data can help communities enroll eligible households in programs designed to assist them and can help communities qualify for grants from the Community Development Block Grant, HOME Investment Partnership Program, Emergency Solutions Grants, Housing Opportunities for Persons With AIDS, and other programs.

Disability asked since 1830—Con.

Provide Health Care to Children and Families

Knowing the disability status of people in families in combination with other information, such as household income, health insurance status, and poverty status, can help communities enroll eligible families in programs designed to assist them. For example, disability data are used to target efforts to enroll eligible people in Marketplace, Medicaid, and the Children's Health Insurance Program (CHIP). Disability data are also used to ensure that Marketplace, Medicare, Medicaid, and CHIP programs are adequately serving these families.

Ensure Equal Opportunity

Knowing the disability status of people in the community in combination with information about housing, voting, employment, and education, helps governments and communities enforce laws, regulations, and policies against discrimination based on disability status. For example, disability data are used to evaluate whether there are health care or public health program disparities based on disability status (Developmental Disabilities Assistance and Bill of Rights Act of 2000).

Provide Assistance to People With Disabilities

Knowing how many people in a community over a certain age have a disability helps local officials provide programs and services to older adults that enable them to remain living safely in their homes and communities (Older Americans Act). Disability status data are also used in programs that provide services and assistance to people with a disability, such as financial assistance with utilities (Low Income Home Energy Assistance Program).

Understand Changes

Knowing whether people with disabilities have the same opportunities in education, employment, voting, homeownership, and many other areas is of interest to researchers, advocacy groups, and policymakers. Communities also need to understand changes in the needs and geographic concentrations of people with disabilities to ensure that they can meet the community's needs during weather events, disasters, and public health emergencies.

Fertility asked since 1890.

25 **In the PAST 12 MONTHS, has this person given birth to any children?**

- ☐ Yes
- ☐ No

A QUESTION ABOUT WHETHER A WOMAN HAD A BABY IN THE LAST YEAR IS USED TO CREATE STATISTICS ABOUT FERTILITY.

Fertility data are used in planning government programs and adjusting other important data, such as the size of the population eligible for different services, as new people are born. These statistics can also be used to project the future size of the population and to understand more about growing families.

FERTILITY DATA HELP COMMUNITIES:

Provide Health Care to Children and Families

Knowing the number of women with a recent birth in combination with other information, such as marital status, labor force status, household income, health insurance status, and poverty status, can help communities understand changes in the demand for health care. For example, knowing how many American Indian babies are born can help communities, tribes, and the federal government estimate the demand for health care through the Indian Health Service.

Understand Changing Households

Knowing the characteristics of women who are giving birth, including where in the country they live, is important to understand the relationships among different development patterns, including housing and travel information, public health, and pollution.

Though local vital statistics offices typically have a count of births per year, fertility data are able to provide federal program planners, policymakers, and researchers with additional statistics about the age, education, and employment of parents in households welcoming children, and other important information about the homes (age, size, etc.) and households (income, language spoken, etc.) for a more complete picture of families.

State and local agencies can use these statistics in combination with other information about new mothers, such as education and income, to understand future needs for the local education system and health services.

This page is intentionally blank.

Grandparent Caregivers asked since 2000.

26 **a. Does this person have any of his/her own grandchildren under the age of 18 living in this house or apartment?**

- ☐ Yes
- ☐ No → *SKIP to question 27*

b. Is this grandparent currently responsible for most of the basic needs of any grandchildren under the age of 18 who live in this house or apartment?

- ☐ Yes
- ☐ No → *SKIP to question 27*

c. How long has this grandparent been responsible for these grandchildren?
If the grandparent is financially responsible for more than one grandchild, answer the question for the grandchild for whom the grandparent has been responsible for the longest period of time.

- ☐ Less than 6 months
- ☐ 6 to 11 months
- ☐ 1 or 2 years
- ☐ 3 or 4 years
- ☐ 5 or more years

QUESTIONS ABOUT WHETHER A PERSON IS THE PRIMARY CAREGIVER FOR HIS/HER GRANDCHILDREN AND HOW LONG HE/SHE HAS CARED FOR HIS/HER GRANDCHILDREN ARE USED TO CREATE STATISTICS ABOUT GRANDPARENT CAREGIVERS.

Grandparent caregiver data help federal agencies understand the special provisions needed for federal programs designed to assist families, as older Americans are often in different financial, housing, and health circumstances than those of other ages. These data are also used to measure the effects of policies and programs that focus on the well-being of families, including tax policies and financial assistance programs.

GRANDPARENT CAREGIVER DATA HELP COMMUNITIES:

Provide Assistance to Families

Knowing more about families, particularly those where grandparents care for grandchildren, along with data about the ages of children, household income, disability, and poverty status can help communities enroll eligible families in programs designed to assist them, such as the Children's Health Insurance Program, and can help communities qualify for grants to fund these programs. These data are also used to evaluate programs like Temporary Assistance for Needy Families.

Provide Assistance to Older Americans

Knowing how many people in a community are over a certain age, including whether older Americans are caring for grandchildren, helps local officials fund programs and services targeted to reach older adults with the greatest economic and social needs (Older Americans Act).

Understand Changing Households

Knowing more about how often grandparents are responsible for the basic care for grandchildren and how long they have been responsible in combination with information about age, presence of children, income, etc., can help communities understand if available housing and services are meeting residents' needs.

This page is intentionally blank.

Health Insurance Coverage asked since 2008.
Health Insurance Premiums and Subsidies starting in 2019.

16 **Is this person CURRENTLY covered by any of the following types of health insurance or health coverage plans?** *Mark "Yes" or "No" for EACH type of coverage in items a – h.*

	Yes	No
a. Insurance through a current or former employer or union (of this person or another family member)	☐	☐
b. Insurance purchased directly from an insurance company (by this person or another family member)	☐	☐
c. Medicare, for people 65 and older, or people with certain disabilities	☐	☐
d. Medicaid, Medical Assistance, or any kind of government-assistance plan for those with low incomes or a disability	☐	☐
e. TRICARE or other military health care	☐	☐
f. VA (including those who have enrolled for VA health care)	☐	☐
g. Indian Health Service	☐	☐
h. Any other type of health insurance or health coverage plan – *Specify*	☐	☐

17 **a. Is there a premium for this plan?** *A premium is a fixed amount of money paid on a regular basis for health coverage. It does not include copays, deductibles, or other expenses such as prescription costs.*

- ☐ Yes
- ☐ No ➔ *SKIP to question 18a*

b. Does this person or another family member receive a tax credit or subsidy based on family income to help pay the premium?

- ☐ Yes
- ☐ No

QUESTIONS ABOUT THE SOURCES OF A PERSON'S HEALTH INSURANCE ARE USED TO CREATE STATISTICS ABOUT THE PERCENTAGE OF PEOPLE COVERED BY HEALTH INSURANCE AND THE SOURCES OF HEALTH INSURANCE.

Health insurance data are used in planning government programs, determining eligibility criteria, and encouraging eligible people to participate in health insurance programs.

HEALTH INSURANCE DATA HELP COMMUNITIES:

Provide Assistance to Children and Families

Knowing the health insurance coverage status in combination with other information, such as number and age of children in families, household income, and poverty status, can help communities enroll eligible families in programs designed to assist them. For example, health insurance coverage status and age data are used to target efforts to enroll eligible people in Marketplace, Medicaid, and the Children's Health Insurance Program (CHIP). Health Insurance data are also used to ensure that Marketplace, Medicare, Medicaid, and CHIP programs are improving health outcomes for families.

Health Insurance Coverage asked since 2008—Con.
Health Insurance Premiums and Subsidies starting in 2019—Con.

Provide Health Care for Veterans

Knowing the number and characteristics of veterans eligible to use Department of Veterans Affairs health care, compared to those currently using services, can help communities and the federal government estimate the future demand for health care services and facilities for veterans.

Provide Health Care for American Indians

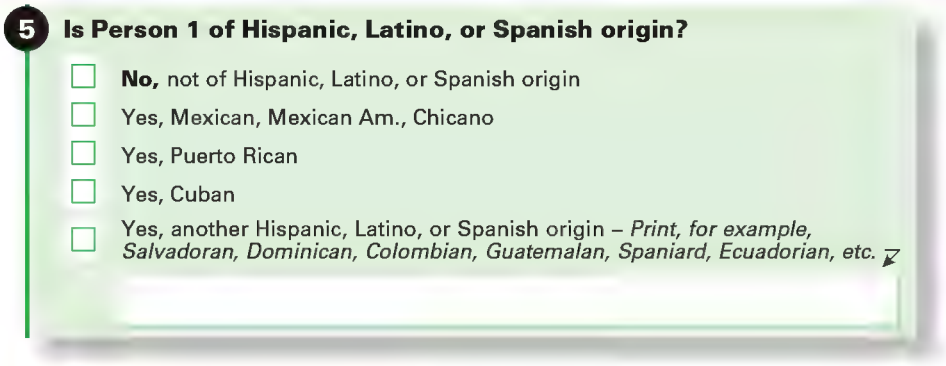
Knowing the health insurance coverage status of American Indians can help communities, tribes, and the federal government estimate the demand for health care through the Indian Health Service.

Understand Changes

Knowing the health insurance coverage status of people in a community helps planners identify gaps in community services, plan programs that address those gaps, and qualify for funding for those programs.

Knowing more about changes in health insurance coverage rates and the characteristics of people who have or do not have health insurance is also of interest to researchers, advocacy groups, and policymakers. For example, State Councils on Developmental Disabilities use health insurance coverage data in their comprehensive reviews and analyses of the unmet needs of people with developmental disabilities.

Hispanic Origin asked since 1970.



A QUESTION ABOUT WHETHER A PERSON IS OF HISPANIC, LATINO, OR SPANISH ORIGIN IS USED TO CREATE STATISTICS ABOUT THIS ETHNIC GROUP.

These data are required for federal and state programs and are factors in the basic research behind numerous policies, particularly for civil rights. Hispanic origin data are used in planning and funding government programs that provide funds or services for specific groups. These data are also used to evaluate government programs and policies to ensure they fairly and equitably serve the needs of the Hispanic population. They are also used to monitor compliance with antidiscrimination laws, regulations, and policies.

Though many respondents expect to see a Hispanic, Latino, or Spanish category on the race question, this question is asked separately because people of Hispanic origin may be of any race(s) in accordance with the 1997 Office of Management and Budget standards on race and ethnicity.

HISPANIC ORIGIN DATA HELP COMMUNITIES:

Ensure Equal Opportunity

Knowing about the Hispanic or Latino origin of community members in combination with information about housing, voting, language, employment, and education, to help governments and communities enforce antidiscrimination laws, regulations, and policies. For example, data on the Hispanic population are used to:

- Establish and evaluate the guidelines for federal affirmative action plans under the Federal Equal Opportunity Recruitment Program.
- Monitor compliance with the Voting Rights Act and enforce bilingual election requirements.
- Monitor and enforce equal employment opportunities under the Civil Rights Act of 1964.
- Identify segments of the population who may not be getting needed medical services under the Public Health Service Act.
- Allocate funds to school districts for bilingual services under the Bilingual Education Act.

Hispanic Origin asked since 1970—Con.

Understand Changes

Researchers, advocacy groups, and policymakers are interested in knowing if people of Hispanic and non-Hispanic origin have the same opportunities in education, employment, voting, and homeownership. The National Science Foundation uses these data to provide information on the Hispanic population in the science and engineering workforce. Several federal agencies use these data to investigate whether housing or transportation improvements have unintended consequences for the Hispanic population. Data on the Hispanic population are used with age and language data to address language and cultural diversity needs in health care plans for the older population.

Home Heating Fuel asked since 1940.

13 **Which FUEL is used MOST for heating this house, apartment, or mobile home?**

- ☐ Gas: from underground pipes serving the neighborhood
- ☐ Gas: bottled, tank, or LP
- ☐ Electricity
- ☐ Fuel oil, kerosene, etc.
- ☐ Coal or coke
- ☐ Wood
- ☐ Solar energy
- ☐ Other fuel
- ☐ No fuel used

QUESTIONS ABOUT HOME HEATING FUEL ARE USED TO CREATE DATA ABOUT HOME ENERGY USE.

These data are used in government programs that analyze community air quality and energy needs. Federal agencies use these statistics to forecast future energy demand, analyze the fuels available to community residents, and plan and fund programs that help low-income residents afford to heat their homes.

HOME HEATING FUEL DATA HELP COMMUNITIES:

Provide Assistance With Utilities

Knowing which fuel is used to heat homes in combination with the cost of those fuels and the characteristics of the low-income households that need assistance with their utilities, helps communities enroll eligible households in assistance programs like the Low Income Home Energy Assistance Program and qualify for grants to fund assistance. These data are also used to evaluate whether these programs benefit eligible households.

Estimate Future Energy Demand

Knowing the current users of certain heating systems and the kinds of systems used in new homes helps communities predict future demand for fuels and the future costs of systems in use in a community. For example, the Department of Energy uses these data to project demand over the next 30 years, assessing the energy needs of the U.S. economy in a domestic and international context.

Measure Environmental Impacts

Communities with older heating systems may have lower air quality at times when they are in high use. Home heating fuel data are used to develop an inventory of the national aggregate emissions of each greenhouse gas and to research and report on the relationships among different development patterns (including housing and travel information) and public health and pollution (Clean Air Act, Clean Water Act).

This page is intentionally blank.

Home Value and Rent asked since 1940.

18 **a. What is the monthly rent for this house, apartment, or mobile home?**

Monthly amount – *Dollars*

$, .00

b. Does the monthly rent include any meals?

☐ Yes
☐ No

19 **About how much do you think this house and lot, apartment, or mobile home (and lot, if owned) would sell for if it were for sale?**

Amount – *Dollars*

$, , .00

QUESTIONS ABOUT THE MONTHLY RENT AMOUNT OR HOW MUCH THE HOME AND PROPERTY ARE WORTH ARE USED TO PRODUCE STATISTICS ABOUT RENT AND HOME VALUE.

These data are used in government programs that analyze whether adequate housing is affordable for residents and provide and fund housing assistance programs. These statistics are also used to enforce laws, regulations, and policies designed to eliminate discrimination in private-market housing, government programs, and in society.

HOME VALUE AND RENT DATA HELP COMMUNITIES:

Provide Adequate Housing

Knowing the different types of households in a community (single people, couples, families, roommates, etc.) helps communities understand whether available housing meets the needs of residents. Housing costs in combination with relationship and combined income of all people in a household helps communities understand whether housing is affordable.

When rental housing is not affordable, rent data are used to identify rental distribution of housing units (the standard cost of different types of housing in different areas of the country) and to determine Fair Market Rents, which the Department of Housing and Urban Development uses to determine the amount of tenant subsidies in housing assistance programs.

When housing is not sufficient or not affordable, housing cost data can help communities enroll eligible households in programs designed to assist them and can help communities qualify for grants from the Community Development Block Grant, HOME Investment Partnership Program, Emergency Solutions Grants, Housing Opportunities for Persons With AIDS, and other programs.

Plan Community Development

Knowing how the balance of rented homes, mortgaged homes, and owned homes changes over time can help communities understand changes in local housing markets and identify opportunities to improve tax, assistance, and zoning policies.

Ensure Equal Opportunity

Knowing more about people who rent and people who own homes in the community in combination with age, sex, race, Hispanic origin, disability, and other data, helps government and communities enforce laws, such as the 1968 Fair Housing Act designed to eliminate discrimination in housing.

This page is intentionally blank.

Income asked since 1940.

43 INCOME IN THE PAST 12 MONTHS

Mark (X) the "Yes" box for each type of income this person received, and give your best estimate of the TOTAL AMOUNT during the PAST 12 MONTHS. (NOTE: The "past 12 months" is the period from today's date one year ago up through today.)

Mark (X) the "No" box to show types of income NOT received.

If net income was a loss, mark the "Loss" box to the right of the dollar amount.

For income received jointly, report the appropriate share for each person – or, if that's not possible, report the whole amount for only one person and mark the "No" box for the other person.

a. Wages, salary, commissions, bonuses, or tips from all jobs. *Report amount before deductions for taxes, bonds, dues, or other items.*

☐ Yes → $ .00
☐ No
TOTAL AMOUNT for past 12 months

b. Self-employment income from own nonfarm businesses or farm businesses, including proprietorships and partnerships. *Report NET income after business expenses.*

☐ Yes → $ .00 ☐ Loss
☐ No
TOTAL AMOUNT for past 12 months

c. Interest, dividends, net rental income, royalty income, or income from estates and trusts. *Report even small amounts credited to an account.*

☐ Yes → $ .00 ☐ Loss
☐ No
TOTAL AMOUNT for past 12 months

d. Social Security or Railroad Retirement.

☐ Yes → $ .00
☐ No
TOTAL AMOUNT for past 12 months

e. Supplemental Security Income (SSI).

☐ Yes → $ .00
☐ No
TOTAL AMOUNT for past 12 months

f. Any public assistance or welfare payments from the state or local welfare office.

☐ Yes → $ .00
☐ No
TOTAL AMOUNT for past 12 months

g. Retirement income, pensions, survivor or disability income. *Include income from a previous employer or union, or any regular withdrawals or distributions from IRA, Roth IRA, 401(k), 403(b) or other accounts specifically designed for retirement. Do not include Social Security.*

☐ Yes → $ .00
☐ No
TOTAL AMOUNT for past 12 months

h. Any other sources of income received regularly such as Veterans' (VA) payments, unemployment compensation, child support or alimony. *Do NOT include lump sum payments such as money from an inheritance or the sale of a home.*

☐ Yes → $ .00
☐ No
TOTAL AMOUNT for past 12 months

44 What was this person's total income during the PAST 12 MONTHS? *Add entries in questions 43a to 43h; subtract any losses. If net income was a loss, enter the amount and mark (X) the "Loss" box next to the dollar amount.*

☐ None OR $ .00 ☐ Loss
TOTAL AMOUNT for past 12 months

Income asked since 1940—Con.

QUESTIONS ABOUT THE FUNDS A PERSON RECEIVES FROM VARIOUS SOURCES ARE USED TO CREATE STATISTICS ABOUT INCOME, ASSISTANCE, EARNINGS, AND POVERTY STATUS.

Income data are used in planning and funding government programs that provide economic assistance for populations in need and measure the economic well-being of the nation. Income and poverty estimates are often part of allocation formulas that determine how food, health care, job training, housing, and other assistance are distributed.

INCOME DATA HELP COMMUNITIES:

Provide Adequate Housing

Knowing the combined income of all people in a household in combination with housing costs helps communities understand whether housing is affordable for residents. When housing is not sufficient or not affordable, income data can help communities enroll eligible households in programs designed to assist them and can help communities qualify for grants from the Community Development Block Grant, HOME Investment Partnership Program, Emergency Solutions Grant, Housing Opportunities for Persons With AIDS, and other programs.

Provide Assistance to Older Americans

Knowing how many older people in a community are living in poverty in combination with other information, such as age and disability status of other family members, can help communities ensure these residents receive appropriate assistance, such as financial assistance with utilities (Low Income Home Energy Assistance Program).

Provide Assistance to Children and Families

Knowing household income in combination with other information, such as the number and age of children in families, health insurance status, and poverty status, can help communities enroll eligible families in programs designed to assist them. For example, income data are used to identify eligibility and provide funding in programs like Medicaid, the Child and Adult Care Food Program, and Head Start.

Educate Children and Adults

Knowing how many children and adults depend on services through schools helps school districts make long-term building, staffing, and funding decisions. Household income and family composition determine poverty status, which is used along with school enrollment, information on disability status, and language spoken at home, to help schools understand the needs of their students and qualify for grants that help fund programs for students with needs for additional services or assistance, including free/reduced price school lunches (Elementary and Secondary Education Act of 1965).

Plan Community Development

Knowing more about the financial situation of residents, including income, employment, and housing costs, can help communities qualify for loan and grant programs designed to stimulate economic recovery, improve housing, run job-training programs, and define areas as empowerment or enterprise zones.

Industry asked since 1820.[1]
Occupation asked since 1850.
Class of Worker asked since 1910.

42 **DESCRIPTION OF EMPLOYMENT**

The next series of questions is about the type of employment this person had last week.

If this person had more than one job, describe the one at which the most hours were worked. If this person did not work last week, describe the most recent employment in the past five years.

a. Which one of the following best describes this person's employment last week or the most recent employment in the past 5 years? *Mark (X) ONE box.*

PRIVATE SECTOR EMPLOYEE

- ☐ **For-profit** company or organization
- ☐ **Non-profit** organization (including tax-exempt and charitable organizations)

GOVERNMENT EMPLOYEE

- ☐ **Local government** (for example: city or county school district)
- ☐ **State government** (including state colleges/universities)
- ☐ **Active duty** U.S. Armed Forces or Commissioned Corps
- ☐ **Federal government** civilian employee

SELF-EMPLOYED OR OTHER

- ☐ **Owner of non-incorporated** business, professional practice, or farm
- ☐ **Owner of incorporated** business, professional practice, or farm
- ☐ Worked **without pay** in a **for-profit** family business or farm for 15 hours or more per week

b. What was the name of this person's employer, business, agency, or branch of the Armed Forces?

c. What kind of business or industry was this? *Include the main activity, product, or service provided at the location where employed. (For example: elementary school, residential construction)*

d. Was this mainly – *Mark (X) ONE box.*

- ☐ manufacturing?
- ☐ wholesale trade?
- ☐ retail trade?
- ☐ other (agriculture, construction, service, government, etc.)?

e. What was this person's main occupation? *(For example: 4th grade teacher, entry-level plumber)*

f. Describe this person's most important activities or duties. *(For example: instruct and evaluate students and create lesson plans, assemble and install pipe sections and review building plans for work details)*

INCOME IN THE PAST 12 MONTHS

QUESTIONS ABOUT A PERSON'S EMPLOYER, THE KIND OF BUSINESS OR INDUSTRY OF THAT EMPLOYER, THE KIND OF WORK A PERSON DOES, AND THAT PERSON'S MOST IMPORTANT ACTIVITIES ARE USED TO PRODUCE INDUSTRY, OCCUPATION, AND CLASS OF WORKER STATISTICS.

These data are used to provide information about the labor force to evaluate government programs and policies to ensure they fairly and equitably serve the needs of all groups, and to enforce laws, regulations, and policies against discrimination in society.

[1] Industry asked in 1820, 1840, and 1910 to present.

Industry asked since 1820—Con.
Occupation asked since 1850—Con.
Class of Worker asked since 1910—Con.

INDUSTRY, OCCUPATION, AND CLASS OF WORKER DATA HELP COMMUNITIES:

Provide Employment Opportunities

Knowing whether programs designed to employ specific groups, such as people with disabilities or veterans, are succeeding is important to employers, federal agencies, and federal government contractors (Vietnam Era Veterans' Readjustment Assistance Act, Rehabilitation Act of 1973). Industry, occupation, and class of worker data provide additional detail about the jobs and careers pursued by people participating in these programs.

State and local agencies use these statistics to identify labor surplus areas (areas with people available for hiring and training), plan workforce development programs including job fairs and training programs, and promote business opportunities.

Ensure Equal Employment Opportunity

Knowing more about people who are employed or looking for work in combination with educational attainment, age, sex, race, Hispanic origin, disability status, veteran status, and other data, helps governments and communities enforce civil rights laws against employment discrimination. For example, these data are used to enforce nondiscrimination in employment by federal agencies, private employers, employment agencies, and labor organizations (Civil Rights Act of 1964).

Understand Changes

Knowing the characteristics of growing or declining industries and occupations is an important part of estimating changes in the economy. Labor force estimates are used in funding decisions; to ensure surveys are accurate, including surveys that provide official labor market estimates; and to understand change in other data (Wagner-Peyser Act and Workforce Investment Act).

Class of worker data, in particular, are used by the National Institute of Food and Agriculture to understand changes in farm workers and agriculture.

Labor Force Status asked since 1890.

30 **a. LAST WEEK, did this person work for pay at a job (or business)?**

- ☐ Yes → *SKIP to question 31*
- ☐ No – Did not work (or retired)

b. LAST WEEK, did this person do ANY work for pay, even for as little as one hour?

- ☐ Yes
- ☐ No → *SKIP to question 36a*

36 **a. LAST WEEK, was this person on layoff from a job?**

- ☐ Yes → *SKIP to question 36c*
- ☐ No

b. LAST WEEK, was this person TEMPORARILY absent from a job or business?

- ☐ Yes, on vacation, temporary illness, maternity leave, other family/personal reasons, bad weather, etc. → *SKIP to question 39*
- ☐ No → *SKIP to question 37*

36 **c. Has this person been informed that he or she will be recalled to work within the next 6 months OR been given a date to return to work?**

- ☐ Yes → *SKIP to question 38*
- ☐ No

37 **During the LAST 4 WEEKS, has this person been ACTIVELY looking for work?**

- ☐ Yes
- ☐ No → *SKIP to question 39*

38 **LAST WEEK, could this person have started a job if offered one, or returned to work if recalled?**

- ☐ Yes, could have gone to work
- ☐ No, because of own temporary illness
- ☐ No, because of all other reasons (in school, etc.)

QUESTIONS ABOUT WHETHER A PERSON WORKED LAST WEEK AND, IF THE ANSWER IS NO, WHETHER HE/SHE WAS LOOKING FOR WORK, WHETHER HE/SHE PLANS TO RETURN TO WORK, AND HOW MUCH THEY WORKED IN THE PAST YEAR ARE USED TO PRODUCE STATISTICS ABOUT THE LABOR FORCE, INCLUDING UNEMPLOYMENT STATISTICS.

Labor force data are used in planning and funding government programs that provide unemployment assistance and services. These data are also used to evaluate other government programs and policies to ensure they fairly and equitably serve the needs of all groups, and to enforce laws, regulations, and policies against discrimination in society.

Labor Force Status asked since 1890—Con.

LABOR FORCE DATA HELP COMMUNITIES:

Provide Employment Opportunities

Knowing whether programs designed to employ specific groups, such as people with disabilities or veterans, are succeeding is important to employers, federal agencies, and federal government contractors (Vietnam Era Veterans' Readjustment Assistance Act, Rehabilitation Act of 1973).

State and local agencies use these statistics to identify labor surplus areas (areas with people available for hiring and training), plan workforce development programs, including job fairs and training programs, and to promote business opportunities.

Ensure Equal Opportunity

Knowing more about people who are employed or looking for work in combination with age, sex, race, Hispanic origin, disability status, veteran status, and other data, helps governments and communities enforce laws, regulations, and policies against discrimination in employment. For example, labor force data are used to enforce nondiscrimination in employment by federal agencies, private employers, employment agencies, and labor organizations (Civil Rights Act of 1964).

Understand Changes

Knowing the characteristics of people who are working or looking for work is an important part of estimating changes in the economy. Labor force estimates are used in funding decisions; to ensure surveys are accurate, including surveys that provide official labor market estimates; and to understand change in other data (Wagner-Peyser Act and Workforce Investment Act).

Language Spoken at Home asked since 1890.[1]

14 **a. Does this person speak a language other than English at home?**

- ☐ Yes
- ☐ No ➔ *SKIP to question 15a*

b. What is this language?

For example: Korean, Italian, Spanish, Vietnamese

c. How well does this person speak English?

- ☐ Very well
- ☐ Well
- ☐ Not well
- ☐ Not at all

QUESTIONS ABOUT WHETHER A PERSON SPEAKS A LANGUAGE OTHER THAN ENGLISH AT HOME, WHAT LANGUAGE HE/SHE SPEAKS, AND HOW WELL HE/SHE SPEAKS ENGLISH ARE USED TO CREATE STATISTICS ABOUT LANGUAGE AND ABILITY TO SPEAK ENGLISH.

Language data are used in planning government programs for adults and children who do not speak English well. These data are also used to ensure that information about public health, law, regulations, voting, and safety is communicated in languages that community members understand.

[1] Language spoken at home was not asked in 1950.

LANGUAGE SPOKEN AT HOME DATA HELP COMMUNITIES:

Educate Children

Knowing how many children and youth with limited English-speaking abilities depend on services through schools helps school districts make long-term staffing and funding decisions. Language spoken at home in combination with other information, such as disability status, school enrollment, and poverty status, helps schools understand the needs of their students and qualify for grants that help fund programs for those students (Elementary and Secondary Education Act of 1965).

Ensure Equal Opportunity

Knowing the languages spoken by people in the community in combination with information about housing, voting, employment, and education, helps the government and communities enforce laws, regulations, and policies against discrimination based on national origin. For example, language data are used to support the enforcement responsibilities under the Voting Rights Act to investigate differences in voter participation rates and to enforce laws and policies related to bilingual requirements.

Knowing languages spoken in a community also helps federal agencies identify needs for services for people with limited English proficiency under Executive Order 13166.

Understand Changes

Knowing whether people who speak languages other than English have the same opportunities in education, employment, voting, homeownership, and many other areas is of interest to researchers, advocacy groups, and policymakers. For example, language data are used with age and ancestry data to address language and cultural diversity needs in health care plans for the older population.

This page is intentionally blank.

Marital Status asked since 1880.
Marital History asked since 1850.

21 **What is this person's marital status?**

- ☐ Now married
- ☐ Widowed
- ☐ Divorced
- ☐ Separated
- ☐ Never married ➔ *SKIP to* **J**

22 **In the PAST 12 MONTHS did this person get –**

	Yes	No
a. Married?	☐	☐
b. Widowed?	☐	☐
c. Divorced?	☐	☐

23 **How many times has this person been married?**

- ☐ Once
- ☐ Two times
- ☐ Three or more times

24 **In what year did this person last get married?**

Year

QUESTIONS ABOUT WHETHER A PERSON IS CURRENTLY MARRIED, WIDOWED, DIVORCED, SEPARATED, OR NEVER MARRIED; WHETHER HIS/HER MARITAL STATUS CHANGED IN THE PAST 12 MONTHS; AND LIFETIME MARRIAGES ARE USED TO CREATE STATISTICS ABOUT CURRENT MARITAL STATUS AND MARITAL HISTORY.

Marital status and marital history data help federal agencies understand marriage trends, forecast future needs of programs that have spousal benefits, and measure the effects of policies and programs that focus on the well-being of families, including tax policies and financial assistance programs.

MARITAL STATUS AND MARITAL HISTORY DATA HELP COMMUNITIES:

Provide Benefits to Spouses and Survivors

Knowing more about how many spouses and ex-spouses may qualify for programs with spousal benefits, including veteran and social security programs, can help federal agencies ensure adequate funding and facilities for these programs and can help communities determine where gaps in benefits and services might exist.

Provide Assistance to Families

Knowing more about families, particularly blended and single-parent families, along with data about the presence of children, labor force status, and poverty status, can help communities enroll eligible families in programs designed to assist them, such as the Children's Health Insurance Program, and can help communities qualify for grants to fund these programs. These data are also used to evaluate programs like Temporary Assistance for Needy Families.

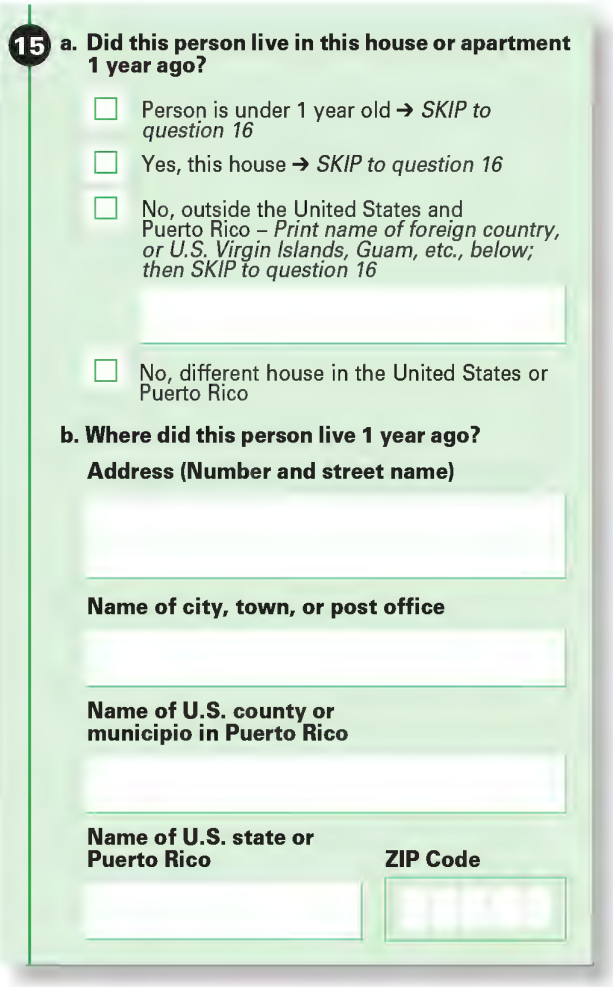
Understand Changing Households

Knowing more about community marriage trends (whether people are marrying later in life, not getting married, or marrying again) in combination with information about age, presence of children, income, etc., can help communities understand if the available housing, job training, rental assistance, and administrative services and programs are meeting residents' needs during their major life changes. These data also help the federal government plan for the future. For example, the Social Security Administration estimates future program needs based on the current relationships of working people.

This page is intentionally blank.

Migration (Previous Residence)/Residence 1 Year Ago

asked since 1930.



15 a. Did this person live in this house or apartment 1 year ago?

- ☐ Person is under 1 year old ➔ *SKIP to question 16*
- ☐ Yes, this house ➔ *SKIP to question 16*
- ☐ No, outside the United States and Puerto Rico – *Print name of foreign country, or U.S. Virgin Islands, Guam, etc., below; then SKIP to question 16*
- ☐ No, different house in the United States or Puerto Rico

b. Where did this person live 1 year ago?

Address (Number and street name)

Name of city, town, or post office

Name of U.S. county or municipio in Puerto Rico

Name of U.S. state or Puerto Rico

ZIP Code

QUESTIONS ABOUT WHETHER A PERSON MOVED IN THE LAST YEAR AND WHERE HE OR SHE LIVED 1 YEAR AGO ARE USED TO CREATE STATISTICS ABOUT WHERE PEOPLE ARE MOVING (TO/FROM FOREIGN COUNTRIES AND WITHIN THE UNITED STATES).

Migration (residence 1 year ago) data are used in planning government programs. The characteristics of people who have moved are also an important part of estimating population changes. These population estimates are used in funding decisions, to ensure surveys are accurate, to understand change in other data, and to produce official international migration estimates.

MIGRATION/RESIDENCE 1 YEAR AGO DATA HELP COMMUNITIES:

Understand Changes

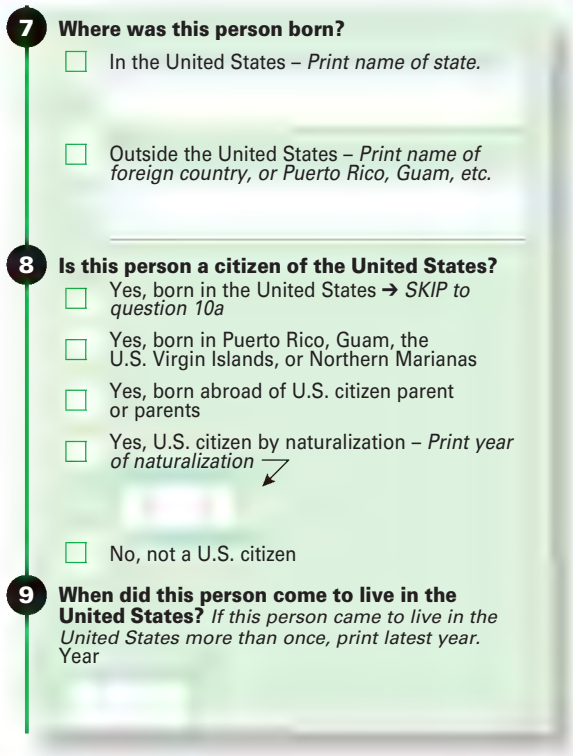
Knowing the characteristics of people who have moved and the patterns of migration (where people move to and from) is an important part of estimating population changes. Population estimates are used in funding decisions, to ensure surveys are accurate, to understand change in other data, and to produce international migration estimates. These data also help agencies assess residential stability and the effects of migration on urban and rural areas.

Knowing where certain populations move to and from helps federal agencies assess the needs of counties with large refugee populations and the effects of immigration on local areas.

Knowing the characteristics of people who live or have lived in certain areas is important to understand the relationships among different development patterns, including housing and travel information, public health, and pollution. These data may also assist state and local agencies in developing programs that attract new residents or employers.

This page is intentionally blank.

Place of Birth asked since 1850.
Citizenship asked since 1820.[1]
Year of Entry asked since 1890.[2]



QUESTIONS ABOUT A PERSON'S PLACE OF BIRTH, CITIZENSHIP, AND YEAR OF ENTRY INTO THE UNITED STATES ARE USED TO CREATE DATA ABOUT CITIZENS, NONCITIZENS, AND THE FOREIGN-BORN POPULATION.

These statistics are essential for agencies and policymakers setting and evaluating immigration policies and laws; seeking to understand the experience of different immigrant groups; and enforcing laws, policies, and regulations against discrimination based on national origin. These statistics are also used to tailor services to accommodate cultural differences.

PLACE OF BIRTH, CITIZENSHIP, AND YEAR OF ENTRY DATA HELP COMMUNITIES:

Ensure Equal Opportunity

Knowing how many people in the community are born in other countries in combination with information about housing, voting, language, employment, and education, helps the government and communities to enforce laws, regulations, and policies against discrimination based on national origin. For example, these data are used to support the enforcement responsibilities under the Voting Rights Act to investigate differences in voter participation rates and to enforce other laws and policies regarding bilingual requirements.

Educate Children

Knowing how many foreign-born children depend on services through schools helps school districts make staffing and funding decisions. Place of birth, citizenship, and year of entry statistics in combination with other information, such as language spoken at home, help schools understand the needs of their students and qualify for grants that help fund programs for those students (Elementary and Secondary Education Act of 1965).

Understand Changes

Knowing whether people of different races or countries of birth have the same opportunities in education, employment, voting, homeownership, and many other areas is of interest to researchers, advocacy groups, and policymakers. These data may also help communities with large refugee populations that qualify for financial assistance (Immigration Nationality Act).

[1] Citizenship asked 1820, 1830, 1870, and 1890 to present.
[2] Year of entry asked 1890–1930 and 1970 to present.

This page is intentionally blank.

Plumbing Facilities asked since 1940.
Kitchen Facilities asked since 1940.
Telephone Service asked since 1960.

7 **Does this house, apartment, or mobile home have –**

	Yes	No
a. hot and cold running water?	☐	☐
b. a bathtub or shower?	☐	☐
c. a sink with a faucet?	☐	☐
d. a stove or range?	☐	☐
e. a refrigerator?	☐	☐

8 **Can you or any member of this household both make and receive phone calls when at this house, apartment, or mobile home?**
Include calls using cell phones, land lines, or other phone devices.

☐ Yes
☐ No

QUESTIONS ABOUT THE PRESENCE OF HOT AND COLD RUNNING WATER, A BATHTUB OR SHOWER, A SINK WITH A FAUCET, A STOVE OR RANGE, A REFRIGERATOR, AND TELEPHONE SERVICE ARE USED TO CREATE DATA ABOUT INDICATORS OF HOUSING QUALITY.

These data are used in planning and funding government programs that identify areas eligible for housing assistance, rehabilitation loans, and other programs that help people access and afford decent, safe, and sanitary housing. Public health officials may also use this information to locate areas in danger of groundwater contamination and waterborne diseases.

PLUMBING FACILITIES, KITCHEN FACILITIES, AND TELEPHONE SERVICE DATA HELP COMMUNITIES:

Provide Adequate Housing

Knowing more about the quality of housing in a community helps communities understand whether available housing meets the needs of residents. When housing is not sufficient or not affordable, data on household facilities can help communities enroll eligible households in programs designed to assist them, and can help communities qualify for grants from the Community Development Block Grant, HOME Investment Partnership Program, Emergency Solutions Grant, Housing Opportunities for Persons With AIDS, and other programs.

Plan Community Development

Knowing how the quality of different types of homes in combination with whether they are occupied or vacant, can help communities identify opportunities to improve tax, assistance, and zoning policies and to reduce tax revenue losses from vacant or abandoned properties. These data may also be useful in identifying types of homes in disaster-prone areas during emergency planning and preparation.

Ensure Residents Can Communicate

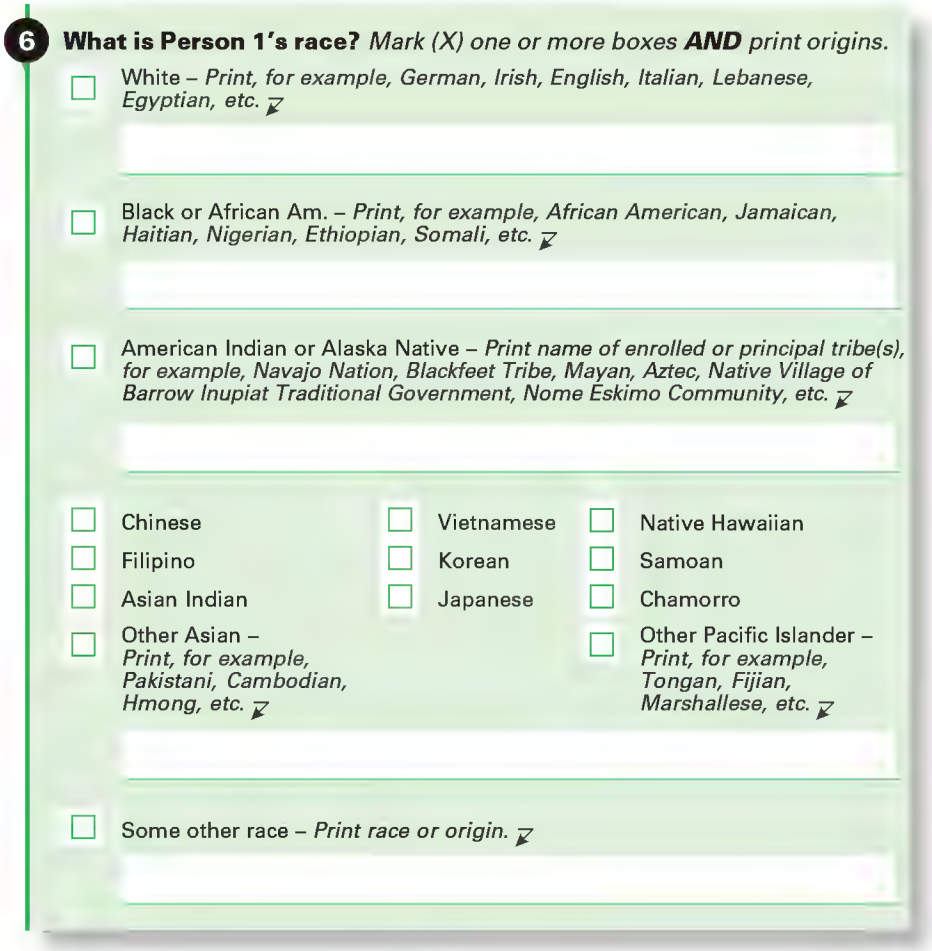
Measuring the extent of telephone service helps communities ensure their residents have universal access to assistance programs, emergency services, and important information.

Measure Environmental Impacts

Substandard plumbing systems may impact the local water supply. Understanding where these systems are concentrated helps communities research their wastewater infrastructure needs and work to improve their systems.

This page is intentionally blank.

Race asked since 1790.



6 **What is Person 1's race?** *Mark (X) one or more boxes* ***AND*** *print origins.*

- ☐ White – *Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc.*
- ☐ Black or African Am. – *Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.*
- ☐ American Indian or Alaska Native – *Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc.*
- ☐ Chinese
- ☐ Filipino
- ☐ Asian Indian
- ☐ Other Asian – *Print, for example, Pakistani, Cambodian, Hmong, etc.*
- ☐ Vietnamese
- ☐ Korean
- ☐ Japanese
- ☐ Native Hawaiian
- ☐ Samoan
- ☐ Chamorro
- ☐ Other Pacific Islander – *Print, for example, Tongan, Fijian, Marshallese, etc.*
- ☐ Some other race – *Print race or origin.*

A QUESTION ABOUT A PERSON'S RACE IS USED TO CREATE STATISTICS ABOUT RACE AND TO PRESENT OTHER ESTIMATES BY RACE GROUPS.

These data are required for federal, state, and tribal programs and are critical factors in the basic research behind numerous policies, particularly for civil rights. Race data are used in planning and funding government programs that provide funds or services for specific groups. These data are also used to evaluate government programs and policies to ensure they fairly and equitably serve the needs of all racial groups and to monitor compliance with antidiscrimination laws, regulations, and policies.

The U.S. Census Bureau collects race data in accordance with the 1997 Office of Management and Budget standards on race and ethnicity. The categories on race are based on self-identification and generally reflect a social definition of race. The categories are not an attempt to define race biologically, anthropologically, or genetically.

RACE DATA HELP COMMUNITIES:

Ensure Equal Opportunity

Knowing the races of community members in combination with information about housing, voting, language, employment, and education,

Race asked since 1790—Con.

helps government and communities enforce antidiscrimination laws, regulations, and policies. For example, race data are used in the following ways:

- Establish and evaluate the guidelines for federal affirmative action plans under the Federal Equal Opportunity Recruitment Program.
- Monitor compliance with the Voting Rights Act and enforce bilingual election requirements.
- Monitor and enforce equal employment opportunities under the Civil Rights Act of 1964.
- Identify segments of the population who may not be getting needed medical services under the Public Health Service Act.
- Allocate funds to school districts for bilingual services under the Bilingual Education Act.

Understand Changes

Knowing if people of different races have the same opportunities in education, employment, voting, homeownership, and many other areas is of interest to researchers, advocacy groups, and policymakers. The National Science Foundation uses data on race to provide information on people of different racial backgrounds in the science and engineering workforce. Several federal agencies use race data to investigate whether housing or transportation improvements have unintended consequences for specific race groups. Data on race are used with age and language data to address language and cultural diversity needs in health care plans for the older population.

Administer Programs for Specific Groups

Knowing how many people are eligible to participate in certain programs helps communities, including tribal governments, ensure that programs are operating as intended. For example, the Indian Housing Block Grant program, Indian Community Development Block Grant program, and Indian Health Service all depend on accurate estimates of American Indians and Alaska Natives. Data for the American Indian and Alaska Native population come from the question about a person's race.

Relationship asked since 1880.

2 **How is this person related to Person 1?** *Mark (X) ONE box.* 2

- ☐ Opposite-sex husband/wife/spouse
- ☐ Opposite-sex unmarried partner
- ☐ Same-sex husband/wife/spouse
- ☐ Same-sex unmarried partner
- ☐ Biological son or daughter
- ☐ Adopted son or daughter
- ☐ Stepson or stepdaughter
- ☐ Brother or sister
- ☐ Father or mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law or daughter-in-law
- ☐ Other relative
- ☐ Roommate or housemate
- ☐ Foster child
- ☐ Other nonrelative

A QUESTION ABOUT THE RELATIONSHIP OF EACH PERSON IN A HOUSEHOLD TO ONE CENTRAL PERSON IS USED TO CREATE ESTIMATES ABOUT FAMILIES, HOUSEHOLDS, AND OTHER GROUPS, AND TO PRESENT OTHER DATA AT A HOUSEHOLD LEVEL.

Relationship data are used in planning and funding government programs that provide funds or services for families, people living or raising children alone, grandparents living with grandchildren, or other households that qualify for additional assistance.

RELATIONSHIP DATA HELP COMMUNITIES:

Provide Adequate Housing

Knowing about the different types of households in a community (single people, couples, families, roommates, etc.) helps communities understand whether available housing meets the needs of residents. Information about the relationships among people in a household, in combination with housing costs and the combined income of all people in a household, helps communities understand whether housing is affordable for residents.

When housing is not sufficient or not affordable, relationship data can help communities enroll eligible households in programs designed to assist them, and can help communities qualify for grants from the Community Development Block Grant, HOME Investment Partnership Program, Emergency Solutions Grant, Housing Opportunities for Persons With AIDS, and other programs.

Provide Assistance to Families

Knowing more about families, such as the ages of children, household income, health insurance status, and poverty status, can help communities enroll eligible families in programs designed to assist them, such as Head Start and the Children's Health Insurance Program, and can help communities qualify for grants to fund these programs. Relationship data are also used to ensure that programs like Temporary Assistance for Needy Families are making a difference for families.

Understand Changing Households

Information about living arrangements and how they are changing, including whether older residents are staying in their homes as they age, whether young people are living with parents or moving in with roommates, and which kinds of households include young children, can help communities plan future programs and services for residents. For example, the Social Security Administration estimates future program needs based on the current relationships of working people.

This page is intentionally blank.

School Enrollment asked since 1850.
Educational Attainment asked since 1940.
Undergraduate Field of Degree asked since 2009.

10 **a. At any time IN THE LAST 3 MONTHS, has this person attended school or college?** *Include only nursery or preschool, kindergarten, elementary school, home school, and schooling which leads to a high school diploma or a college degree.*

- ☐ No, has not attended in the last 3 months → *SKIP to question 11*
- ☐ Yes, public school, public college
- ☐ Yes, private school, private college, home school

b. What grade or level was this person attending? *Mark (X) ONE box.*

- ☐ Nursery school, preschool
- ☐ Kindergarten
- ☐ Grade 1 through 12 – *Specify grade 1 – 12* ☐☐
- ☐ College undergraduate years (freshman to senior)
- ☐ Graduate or professional school beyond a bachelor's degree *(for example: MA or PhD program, or medical or law school)*

11 **What is the highest degree or level of school this person has COMPLETED?** *Mark (X) ONE box. If currently enrolled, mark the previous grade or highest degree received.*

NO SCHOOLING COMPLETED

- ☐ No schooling completed

NURSERY OR PRESCHOOL THROUGH GRADE 12

- ☐ Nursery school
- ☐ Kindergarten
- ☐ Grade 1 through 11 – *Specify grade 1 – 11* ☐☐
- ☐ 12th grade – **NO DIPLOMA**

HIGH SCHOOL GRADUATE

- ☐ Regular high school diploma
- ☐ GED or alternative credential

COLLEGE OR SOME COLLEGE

- ☐ Some college credit, but less than 1 year of college credit
- ☐ 1 or more years of college credit, no degree
- ☐ Associate's degree *(for example: AA, AS)*
- ☐ Bachelor's degree *(for example: BA, BS)*

AFTER BACHELOR'S DEGREE

- ☐ Master's degree *(for example: MA, MS, MEng, MEd, MSW, MBA)*
- ☐ Professional degree beyond a bachelor's degree *(for example: MD, DDS, DVM, LLB, JD)*
- ☐ Doctorate degree *(for example: PhD, EdD)*

12 **This question focuses on this person's BACHELOR'S DEGREE. Please print below the specific major(s) of any BACHELOR'S DEGREES this person has received.** *(For example: chemical engineering, elementary teacher education, organizational psychology)*

School Enrollment asked since 1850—Con.
Educational Attainment asked since 1940—Con.
Undergraduate Field of Degree asked since 2009—Con.

QUESTIONS ABOUT WHETHER A PERSON IS ATTENDING SCHOOL OR COLLEGE, THE HIGHEST LEVEL OF EDUCATION HE/SHE HAS COMPLETED, AND THE FIELD OF ANY COMPLETED UNDERGRADUATE COLLEGE DEGREES ARE USED TO CREATE DATA ABOUT EDUCATION.

These statistics are used to analyze the characteristics and needs of school-aged children and to understand the continuing education needs of adults.

SCHOOL ENROLLMENT, EDUCATIONAL ATTAINMENT, AND UNDERGRADUATE FIELD OF DEGREE DATA HELP COMMUNITIES:

Educate Children and Adults

Knowing how many children and adults depend on services through schools helps school districts make long-term building, staffing, and funding decisions. School enrollment in combination with other information, such as disability status, language spoken at home, and poverty status, helps schools understand the needs of their students and qualify for grants that help fund programs for those students (Elementary and Secondary Education Act of 1965).

Knowing how many adults do not have a high school diploma or equivalent helps schools understand the needs of adult students and qualify for grants that help fund programs for these students (Workforce Investment Act).

Knowing the major fields of study of adults with bachelor's degrees enables efforts to develop the nation's science, technology, engineering, and mathematics labor force (America COMPETES Reauthorization Act of 2010).

Ensure Equal Opportunity

Understanding more about the characteristics of people enrolled or not enrolled in school helps government and communities enforce laws, regulations, and policies against discrimination in education (Civil Rights Act of 1964).

Knowing the educational attainment of workers compared to those seeking employment in combination with age, sex, race, Hispanic origin, disability, and other data, helps enforce nondiscrimination in employment by federal agencies, private employers, employment agencies, and labor organizations (Civil Rights Act of 1964).

Selected Monthly Owner Costs (Cost of Utilities, Condominium and Mobile Home Fees, Taxes, Insurance, and Mortgages).[1]

14 a. LAST MONTH, what was the cost of electricity for this house, apartment, or mobile home?

Last month's cost – *Dollars*

$, .00

OR

☐ Included in rent or condominium fee

☐ No charge or electricity not used

b. LAST MONTH, what was the cost of gas for this house, apartment, or mobile home?

Last month's cost – *Dollars*

$, .00

OR

☐ Included in rent or condominium fee

☐ Included in electricity payment entered above

☐ No charge or gas not used

c. IN THE PAST 12 MONTHS, what was the cost of water and sewer for this house, apartment, or mobile home? *If you have lived here less than 12 months, estimate the cost.*

Past 12 months' cost – *Dollars*

$, .00

OR

☐ Included in rent or condominium fee

☐ No charge

d. IN THE PAST 12 MONTHS, what was the cost of oil, coal, kerosene, wood, etc., for this house, apartment, or mobile home? *If you have lived here less than 12 months, estimate the cost.*

Past 12 months' cost – *Dollars*

$, .00

OR

☐ Included in rent or condominium fee

☐ No charge or these fuels not used

16 Is this house, apartment, or mobile home part of a condominium?

☐ Yes → **What is the monthly condominium fee?** *For renters, answer only if you pay the condominium fee in addition to your rent; otherwise, mark the "None" box.*

Monthly amount – *Dollars*

$, .00

OR

☐ None

☐ No

20 What are the annual real estate taxes on THIS property?

Annual amount – *Dollars*

$, .00

OR

☐ None

21 What is the annual payment for fire, hazard, and flood insurance on THIS property?

Annual amount – *Dollars*

$, .00

OR

☐ None

24 What are the total annual costs for personal property taxes, site rent, registration fees, and license fees on THIS mobile home and its site? *Exclude real estate taxes.*

Annual costs – *Dollars*

$, .00

22 a. Do you or any member of this household have a mortgage, deed of trust, contract to purchase, or similar debt on THIS property?

☐ Yes, mortgage, deed of trust, or similar debt

☐ Yes, contract to purchase

☐ No → *SKIP to question 23a*

b. How much is the regular monthly mortgage payment on THIS property? *Include payment only on FIRST mortgage or contract to purchase.*

Monthly amount – *Dollars*

$, .00

OR

☐ No regular payment required → *SKIP to question 23a*

c. Does the regular monthly mortgage payment include payments for real estate taxes on THIS property?

☐ Yes, taxes included in mortgage payment

☐ No, taxes paid separately or taxes not required

d. Does the regular monthly mortgage payment include payments for fire, hazard, or flood insurance on THIS property?

☐ Yes, insurance included in mortgage payment

☐ No, insurance paid separately or no insurance

23 a. Do you or any member of this household have a second mortgage or a home equity loan on THIS property?

☐ Yes, home equity loan

☐ Yes, second mortgage

☐ Yes, second mortgage and home equity loan

☐ No → *SKIP to* **D**

b. How much is the regular monthly payment on all second or junior mortgages and all home equity loans on THIS property?

Monthly amount – *Dollars*

$, .00

OR

☐ No regular payment required

[1] Cost of utilities asked since 1940, condominium and mobile home fees asked since 1990, taxes asked in 1940 and since 1980, insurance cost asked since 1980, and mortgages cost asked in 1940 and since 1980.

Selected Monthly Owner Costs (Cost of Utilities, Condominium and Mobile Home Fees, Taxes, Insurance, and Mortgages)—Con.

QUESTIONS ABOUT THE USE AND COST OF COMMON UTILITIES, ANY APPLICABLE CONDOMINIUM AND MOBILE HOME FEES, TAXES, MORTGAGES, AND HOME LOANS ARE USED TO PRODUCE STATISTICS ABOUT SELECTED MONTHLY OWNER COSTS.

These data are used in government programs that analyze whether adequate housing is affordable for residents and to provide and fund housing assistance programs. These statistics are also used to enforce laws, regulations, and policies against discrimination in government programs and in society.

SELECTED MONTHLY OWNER COSTS DATA HELP COMMUNITIES:

Provide Adequate Housing

Comparing housing costs to household income (the combined income of everyone in the household) helps communities understand whether housing is affordable for residents.

When housing is not sufficient or not affordable, housing cost data can help communities enroll eligible households in programs designed to assist them, and can help communities qualify for grants from the Community Development Block Grant, HOME Investment Partnerships Program, Emergency Solutions Grants, Housing Opportunities for Persons With AIDS, and other programs.

Plan Community Development

Knowing how housing costs change over time can help communities understand changes in local housing markets and to identify opportunities to improve tax, assistance, and zoning policies.

Ensure Equal Opportunity

Knowing more about the housing costs of people who own homes in the community in combination with age, sex, race, Hispanic origin, disability, and other data about the household residents, helps government and communities enforce laws designed to eliminate discrimination in housing (Fair Housing Act of 1968).

Sex asked since 1790.

3 **What is Person 1's sex?** *Mark (X) ONE box.*

☐ Male ☐ Female

A QUESTION ABOUT THE SEX OF EACH PERSON IS USED TO CREATE STATISTICS ABOUT MALES AND FEMALES AND TO PRESENT OTHER DATA, SUCH AS OCCUPATION, BY SEX.

Data disaggregated by sex are used in planning and funding government programs and in evaluating other government programs and policies to ensure they fairly and equitably serve the needs of males and females. These statistics are also used to enforce laws, regulations, and policies against discrimination in government programs and in society.

DATA DISAGGREGATED BY SEX HELP COMMUNITIES:

Ensure Equal Opportunity

Knowing the distribution of men and women in the community in combination with information about housing, voting, language, employment, and education, helps government and communities enforce laws, regulations, and policies against discrimination on the basis of sex. For example, sex-disaggregated data are used to enforce laws against discrimination based on sex in education programs and activities receiving federal financial assistance (Title IX of the Education Amendments of 1972).

Understand Changes

Knowing whether men and women have the same opportunities in education, employment, voting, homeownership, and many other areas is of interest to researchers, advocacy groups, and policymakers. For example, the National Science Foundation uses data disaggregated by sex to provide information on women in the science and engineering workforce, and several agencies use sex-disaggregated data to investigate whether women, including women who are military veterans, have similar employment opportunities as men.

This page is intentionally blank.

Supplemental Nutrition Assistance Program (SNAP)/ Food Stamps asked since 2005.

15 **IN THE PAST 12 MONTHS, did you or any member of this household receive benefits from the Food Stamp Program or SNAP (the Supplemental Nutrition Assistance Program)?** *Do NOT include WIC, the School Lunch Program, or assistance from food banks.*

- ☐ Yes
- ☐ No

QUESTIONS ABOUT A HOUSEHOLD'S RECEIPT OF FOOD STAMPS/ SUPPLEMENTAL NUTRITION ASSISTANCE PROGRAM (SNAP)[1] ARE USED TO CREATE STATISTICS ABOUT PARTICIPATION IN FOOD ASSISTANCE PROGRAMS.

SNAP data are used in planning and funding government programs that provide food assistance and in evaluating other government programs.

SNAP DATA HELP COMMUNITIES:

Provide Food Assistance to School Children

Knowing more about food assistance program participation in combination with school enrollment, income, and poverty status, can help communities streamline administration of the National School Lunch Program and School Breakfast Program by replacing administrative paperwork with American Community Survey estimates of students eligible for free and reduced-price meals.

Evaluate SNAP

Knowing more about food assistance program participation helps the U.S. Department of Agriculture to evaluate SNAP and award bonuses to communities that administer SNAP funds well.

Understand Changes

State and local agencies use these statistics to assess state food assistance needs and participation rates for eligible families and individuals and to determine gaps in services and programs. Faith-based and other nonprofit organizations use information about food assistance needs to determine where food banks, food kitchens, and other programs could be beneficial and how the needs of their communities can be met with additional resources and services.

[1] In 2008, the food stamp program was renamed SNAP, but the question uses both program names to minimize confusion.

This page is intentionally blank.

Tenure (Owner/Renter) asked since 1890.

17 **Is this house, apartment, or mobile home –** *Mark (X) ONE box.*

- ☐ Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*
- ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
- ☐ Rented?
- ☐ Occupied without payment of rent? ➔ *SKIP to* **C** *on the next page*

A QUESTION ABOUT WHETHER A HOME IS OWNED OR RENTED IS USED TO CREATE DATA ABOUT TENURE, RENTERS, AND HOMEOWNERSHIP.

Tenure is the most basic characteristic to assess housing inventory. Tenure data are used in government programs that analyze whether adequate housing is affordable for residents. Tenure data are also used to provide and fund housing assistance programs. These statistics are also used to enforce laws, regulations, and policies against discrimination in private-market housing, government programs, and in society.

TENURE DATA HELP COMMUNITIES:

Provide Adequate Housing

Knowing the different types of households in a community (single people, couples, families, roommates, etc.) and rates of home rental and ownership helps communities understand whether available housing meets the needs of residents. Data about owners and renters, in combination with housing costs and the combined income of all people in a household, help communities understand whether housing is affordable for residents.

When housing is not sufficient or not affordable, data about owners and renters can help communities enroll eligible households in programs designed to assist them, and can help communities qualify for grants from the Community Development Block Grant, HOME Investment Partnership Program, Emergency Solutions Grants, Housing Opportunities for Persons With AIDS, and other programs.

Plan Community Development

Knowing how the balance of rented homes, mortgaged homes, and homes owned free and clear changes over time can help communities understand changes in local housing markets; identify opportunities to improve tax, assistance, and zoning policies; and reduce tax revenue losses from vacant or abandoned properties. Tenure is also used in formulas that communities use to determine housing assistance funding (Fair Market Rents).

Ensure Equal Opportunity

Knowing the characteristics of people who rent and people who own homes in the community, such as age, sex, race, Hispanic origin, and disability, helps government and communities enforce laws designed to eliminate discrimination in housing (Fair Housing Act of 1968).

Understand Changing Households

Knowing whether older residents are staying in homes as they age or moving into rented homes, and whether young people are staying with parents, renting with roommates, or buying homes, can help governments and communities distribute funds appropriately between homeownership and rental housing programs and services for residents.

This page is intentionally blank.

Units in Structure asked since 1940.
Rooms asked since 1940.
Bedrooms asked since 1960.

1 **Which best describes this building?**
Include all apartments, flats, etc., even if vacant.

- ☐ A mobile home
- ☐ A one-family house detached from any other house
- ☐ A one-family house attached to one or more houses
- ☐ A building with 2 apartments
- ☐ A building with 3 or 4 apartments
- ☐ A building with 5 to 9 apartments
- ☐ A building with 10 to 19 apartments
- ☐ A building with 20 to 49 apartments
- ☐ A building with 50 or more apartments
- ☐ Boat, RV, van, etc.

6 **a. How many separate rooms are in this house, apartment, or mobile home?**
Rooms must be separated by built-in archways or walls that extend out at least 6 inches and go from floor to ceiling. 10

- *INCLUDE bedrooms, kitchens, etc.*
- *EXCLUDE bathrooms, porches, balconies, foyers, halls, or unfinished basements.*

Number of rooms

b. How many of these rooms are bedrooms?
Count as bedrooms those rooms you would list if this house, apartment, or mobile home were for sale or rent. If this is an efficiency/studio apartment, print "0". 11

Number of bedrooms

QUESTIONS ABOUT THE TYPE OF BUILDING, UNITS IN THE STRUCTURE, NUMBER OF ROOMS, AND NUMBER OF BEDROOMS ARE USED TO CREATE DATA ABOUT HOUSING TYPES AND HOUSING DENSITY.

These data are used in government programs that analyze whether adequate housing is available and affordable for residents, and provide and fund housing assistance programs. The number of rooms in combination with the number of people living in a unit provides a ratio of people to rooms, which can be used to measure the extent of overcrowding among our nation's households. These statistics are also used to enforce laws, policies, and regulations against discrimination in government programs and in society.

UNITS IN STRUCTURE, ROOMS, AND BEDROOMS DATA HELP COMMUNITIES:

Provide Adequate Housing

Knowing the different types of housing, and how many people occupy that housing, helps communities understand whether available housing meets the needs of residents. For example, these data are used to measure overcrowding in communities and are used as integral components to set Fair Market Rents for all areas of the country.

Units in Structure asked since 1940—Con.
Rooms asked since 1940—Con.
Bedrooms asked since 1960—Con.

When housing is not sufficient, data can help communities enroll eligible households in programs designed to assist them (such as the Low Income Home Energy Assistance Program), and can help communities qualify for grants from the Community Development Block Grant, HOME Investment Partnerships Program, Emergency Solutions Grants, Housing Opportunities for Persons With AIDS, and other programs.

These data provide benchmark statistics that measure progress toward the Congressional declaration of goals for a national housing policy—a decent home and suitable living environment for every American family.

Plan Community Development

These data are used to identify adequate housing and may be useful in identifying types of structures in disaster-prone areas during emergency planning and preparation.

Vehicles Available asked since 1960.

12 **How many automobiles, vans, and trucks of one-ton capacity or less are kept at home for use by members of this household?**

- ☐ None
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6 or more

A QUESTION ABOUT THE VEHICLES AVAILABLE TO EACH HOUSEHOLD IS USED TO CREATE DATA ABOUT VEHICLE ACCESS.

Vehicle data are used in planning and funding for improvements to road and highway infrastructure, developing transportation plans and services, and understanding how people are traveling in the course of a normal day. These data are also used to evaluate pollution and access to transportation in emergencies.

VEHICLE AVAILABILITY DATA HELP COMMUNITIES:

Improve Transportation

Knowing how many households have access to vehicles, in combination with where people commute to and from, and whether they commute with a personal vehicle helps transportation planners create mass transportation and metropolitan plans that are compliant with various regulations.

Local agencies and organizations use these data to plan programs and services for the disabled population, bicycle commuters, carpools and ride-sharers, and many other groups; and to predict future use of new or updated transportation systems based on their understanding of the current users of various transportation options.

Understand Changes in Vehicle Use

Understanding vehicle availability and use helps communities understand exposure to air pollution and plan programs to help people without vehicles move about the community. Knowing whether people could evacuate using their personal vehicles in an emergency also helps communities plan emergency response.

This page is intentionally blank.

Veteran Status asked since 1890.[1]
Period of Military Service asked since 1890.[1]
Department of Veterans Affairs (VA) Service-Connected Disability Rating asked since 2008.

27 **Has this person ever served on active duty in the U.S. Armed Forces, Reserves, or National Guard?** *Mark (X) ONE box.*

- ☐ Never served in the military ➔ *SKIP to question 30a*
- ☐ Only on active duty for training in the Reserves or National Guard ➔ *SKIP to question 29a*
- ☐ Now on active duty
- ☐ On active duty in the past, but not now

28 **When did this person serve on active duty in the U.S. Armed Forces?** *Mark (X) a box for EACH period in which this person served, even if just for part of the period.*

- ☐ September 2001 or later
- ☐ August 1990 to August 2001 (including Persian Gulf War)
- ☐ May 1975 to July 1990
- ☐ Vietnam era (August 1964 to April 1975)
- ☐ February 1955 to July 1964
- ☐ Korean War (July 1950 to January 1955)
- ☐ January 1947 to June 1950
- ☐ World War II (December 1941 to December 1946)
- ☐ November 1941 or earlier

29 **a. Does this person have a VA service-connected disability rating?**

- ☐ Yes (such as 0%, 10%, 20%, ... , 100%)
- ☐ No ➔ *SKIP to question 30a*

b. What is this person's service-connected disability rating?

- ☐ 0 percent
- ☐ 10 or 20 percent
- ☐ 30 or 40 percent
- ☐ 50 or 60 percent
- ☐ 70 percent or higher

QUESTIONS ABOUT A PERSON'S MILITARY SERVICE AND SERVICE-CONNECTED DISABILITY RATING ARE USED TO CREATE ESTIMATES OF VETERANS AND THEIR NEEDS AT THE COMMUNITY LEVEL.

Data about veterans are used in planning and funding government programs that provide funds or services for veterans and in evaluating other government programs and policies to ensure they fairly and equitably serve the needs of veterans. These statistics are also used to enforce laws, policies, and regulations against discrimination in society. Though the VA maintains veterans' records, these statistics do not provide federal program planners, policymakers, and researchers with additional statistics about all veterans, regardless of whether they use VA services.

VETERAN STATUS, PERIOD OF SERVICE, AND VA SERVICE-CONNECTED DISABILITY RATING DATA HELP COMMUNITIES:

Administer Programs for Veterans

Knowing the numbers and characteristics of veterans eligible for federal programs benefiting veterans, such as the VA Home Loan Guarantee program, the Post-9/11 GI Bill, and job training and hiring preference programs can help communities and the federal government estimate the future demand for these programs and services. These data are also used to evaluate these programs to determine whether they are benefiting veterans as intended.

Provide Health Care for Veterans

Knowing the number of veterans eligible to use VA health care in combination with age, disability, and service-connected disability ratings, can help communities and the federal government estimate

[1] Veteran status and period of service were not asked in 1920.

Veteran Status asked since 1890—Con.
Period of Military Service asked since 1890—Con.
Department of Veterans Affairs (VA) Service-Connected Disability Rating asked since 2008—Con.

the future demand for health care services and facilities. Communities in need of major VA medical facilities throughout the country make a case for new construction projects using these data to estimate the expected usage of new facilities.

Plan End-of-Life Options for Veterans

Knowing where veterans are living toward the end of their lives is important, as the VA estimates the number of nursing home and domiciliary beds needed based on the concentrations of eligible veterans over age 65. These data are also important for the VA National Cemetery Administration, whose goal is to have a VA burial option within 75 miles of a veteran's residence. These data are used to plan construction of new cemeteries near the communities where veterans choose to live.

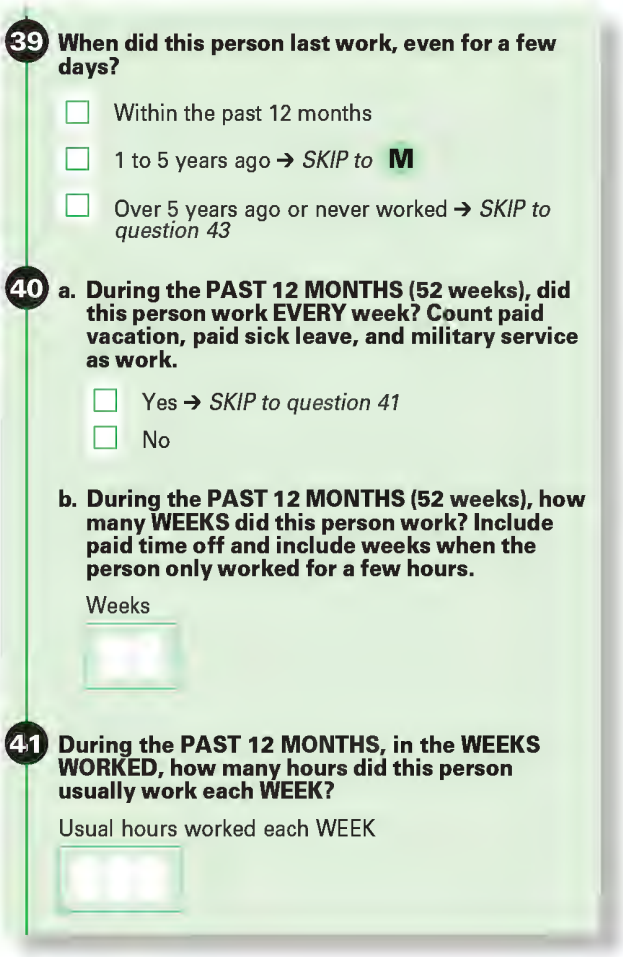
Ensure Equal Opportunity

Knowing the veteran and service-connected disability rating status of people in the community in combination with information about housing, voting, employment, and education, helps government and communities enforce against discrimination based on veteran or disability status.

Understand New Challenges for Veterans

Knowing more about the characteristics of veterans returning to civilian life is also important to combat specific problems they may face. For example, these data are used in research to understand why veteran status is a predictor of homelessness. Such data have been combined with administrative data produced by shelters in an attempt to understand and document which interventions reduce homelessness among veterans.

Work Status Last Year asked since 1880.



39 **When did this person last work, even for a few days?**

- ☐ Within the past 12 months
- ☐ 1 to 5 years ago → *SKIP to* **M**
- ☐ Over 5 years ago or never worked → *SKIP to question 43*

40 **a. During the PAST 12 MONTHS (52 weeks), did this person work EVERY week? Count paid vacation, paid sick leave, and military service as work.**

- ☐ Yes → *SKIP to question 41*
- ☐ No

b. During the PAST 12 MONTHS (52 weeks), how many WEEKS did this person work? Include paid time off and include weeks when the person only worked for a few hours.

Weeks

41 **During the PAST 12 MONTHS, in the WEEKS WORKED, how many hours did this person usually work each WEEK?**

Usual hours worked each WEEK

QUESTIONS ABOUT HOW MANY WEEKS A PERSON WORKED IN THE LAST YEAR AND HOW MANY HOURS HE OR SHE WORKED EACH WEEK ARE USED TO PRODUCE STATISTICS ABOUT FULL-TIME AND PART-TIME WORKERS, AS WELL AS YEAR-ROUND AND SEASONAL WORKERS.

Data on work status last year are used in planning and funding government programs that provide unemployment assistance and services, and to understand trends and difference in wages, benefits, work hours, and seasonal work. These data are also used to evaluate other government programs and policies to ensure they fairly and equitably serve the needs of all groups, and to enforce laws, regulations, and policies against discrimination in society.

WORK STATUS LAST YEAR DATA HELP COMMUNITIES:

Provide Employment Opportunities

Knowing whether programs designed to employ specific groups, such as people with disabilities or veterans, are succeeding is important to employers, federal agencies, and federal government contractors (Vietnam Era Veterans' Readjustment Assistance Act, Rehabilitation Act of 1973).

State and local agencies use these statistics to identify labor surplus areas (areas with people available for hiring and training), plan workforce development programs including job fairs and training programs, and promote business opportunities.

Ensure Equal Opportunity

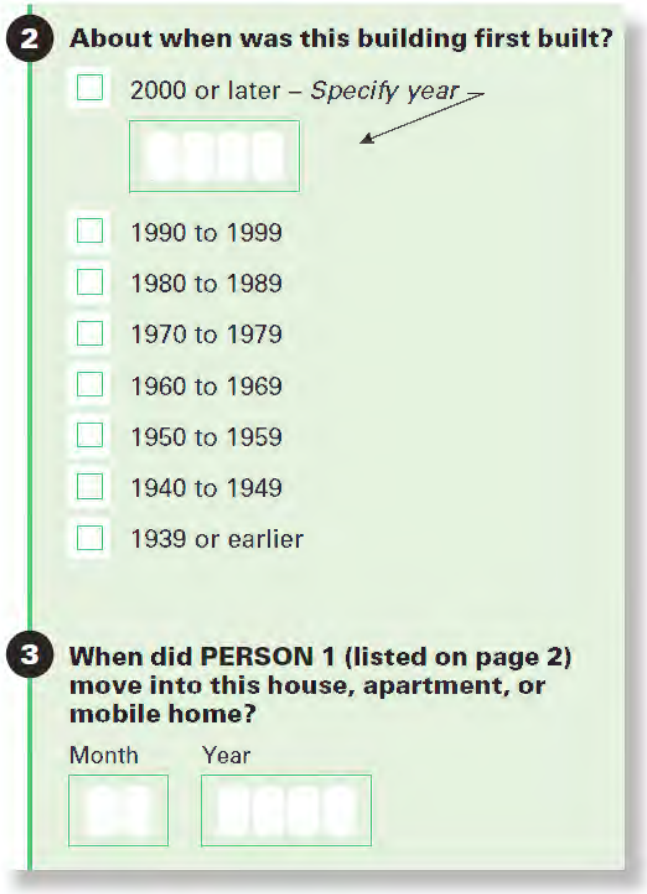
Knowing more about people who are employed or looking for work, in combination with age, sex, race, Hispanic origin, disability status, veteran status, and other data, helps governments and communities enforce laws, policies, and regulations against discrimination in employment. For example, data on work status last year are used to enforce laws against discrimination in employment by federal agencies, private employers, employment agencies, and labor organizations (Civil Rights Act of 1964).

Understand Changes

Knowing the characteristics of people who are working or looking for work is an important part of estimating changes in the economy. Estimates of work status last year are used in funding decisions; to ensure surveys are accurate, including surveys that provide official labor market estimates; and to understand change in other data (Wagner-Peyser Act and Workforce Investment Act).

This page is intentionally blank.

Year Built asked since 1940.
Year Moved In asked since 1960.



QUESTIONS ABOUT WHEN A BUILDING WAS BUILT AND WHEN A PERSON MOVED INTO THAT HOME ARE USED TO CREATE DATA ABOUT HOUSING AGE AND AVAILABILITY.

These data are used in government programs that analyze whether adequate housing is available and affordable for residents, provide and fund housing assistance programs, and measure neighborhood stability.

YEAR BUILT AND YEAR MOVED IN DATA HELP COMMUNITIES:

Provide Adequate Housing

Knowing the ages of housing in a community helps communities understand whether available housing meets the needs of residents.

When housing is not sufficient or older than a certain age, housing data can help communities enroll eligible households in programs designed to assist them (such as the Low Income Home Energy Assistance Program), and can help communities qualify for grants from the Community Development Block Grant, HOME Investment Partnerships Program, Emergency Solutions Grants, Housing Opportunities for Persons With AIDS, and other programs.

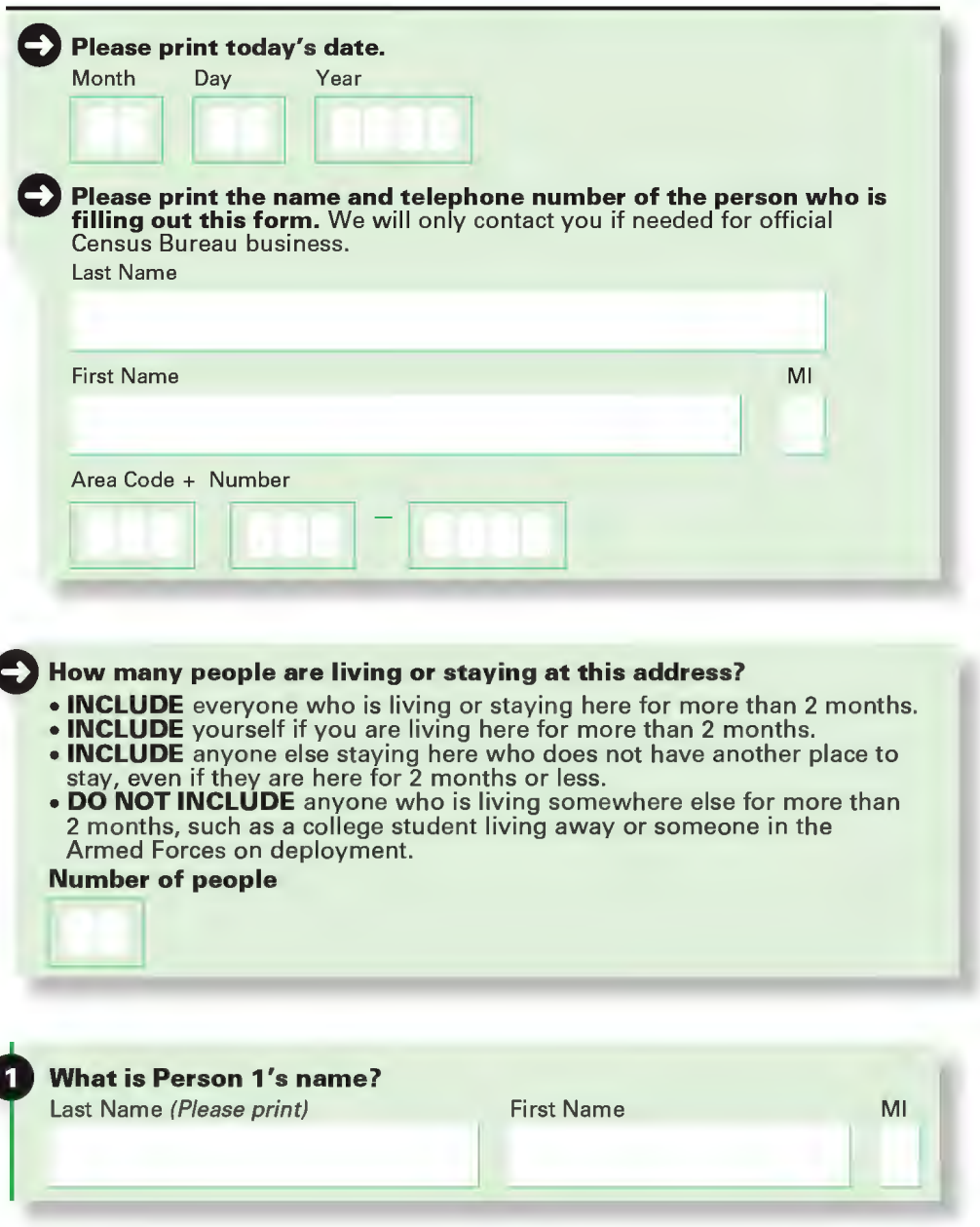
Plan Community Development

Knowing how the balance of different ages of homes in combination with whether they are occupied or vacant, can help communities identify opportunities to improve tax, assistance, and zoning policies and to reduce tax revenue losses from vacant or abandoned properties. These data may also be useful in identifying older structures in disaster-prone areas during emergency planning and preparation.

Knowing more about the age of the housing stock in combination with the financial situation of residents, including income, employment, and housing costs, can help communities qualify for loan and grant programs designed to stimulate economic recovery, improve housing, run job-training programs, and define areas as empowerment or enterprise zones.

This page is intentionally blank.

Operational Questions for use in the American Community Survey.



MEETING FEDERAL NEEDS

Some operational questions will appear on the 2020 ACS that will not result in published estimates. These questions are asked to better administer the data collection process, ensure response accuracy and completeness, and for contacting respondents with incomplete or missing information. Contact information is not part of published estimates and is carefully protected, as mandated by federal law, to protect the personal information of respondents.

This page is intentionally blank.

Appendix A:
Year the Planned Questions Were First Asked in the Decennial Census Program

This page is intentionally blank.

Year the Planned Questions Were First Asked in the Decennial Census Program

Questions Planned for 2020 Census and/or ACS	Year Question First Asked in Decennial Census or ACS	Years Not Asked
Acreage	1960	
Age	1790	
Agricultural Sales	1960	
Ancestry	1980	
Bedrooms	1960	
Citizenship	1820	1840–1860, 1880
Class of Worker	1910	
Commuting (Journey to Work)	1960	
Computer and Internet Use	2013	
Condominium and Mobile Home Fees	1990	
Cost of Utilities	1940	
Department of Veterans Affairs (VA) Service-Connected Disability Rating	2008	
Disability	1830	
Educational Attainment	1940	
Fertility	1890	
Grandparent Caregivers	2000	
Health Insurance Coverage	2008	
Health Insurance Premiums and Subsidies	2019	
Hispanic Origin	1970	
Home Heating Fuel	1940	
Home Value and Rent	1940	
Income	1940	
Industry	1820	1830, 1850–1900
Insurance	1980	
Kitchen Facilities	1940	
Labor Force Status	1890	
Language Spoken at Home	1890	1950
Marital History	1850	
Marital Status	1880	
Migration	1930	
Mortgages	1940	1950–1970
Occupation	1850	
Period of Military Service	1890	1920
Place of Birth	1850	
Plumbing Facilities	1940	
Race	1790	
Relationship	1880	
Rooms	1940	

Year Current Questions Planned First Asked in Decennial Census Program—Con.

Questions Planned for 2020 Census and/or ACS	Year Question First Asked in Decennial Census or ACS	Years Not Asked
School Enrollment	1850	
Sex	1790	
Supplemental Nutrition Assistance Program (SNAP)/ Food Stamps	2005	
Taxes	1940	1950–1970
Telephone Service	1960	
Tenure (Owner/Renter)	1890	
Undergraduate Field of Degree	2009	
Units in Structure	1940	
Vehicles	1960	
Veteran Status	1890	1920
Work Status Last Year	1880	
Year Built	1940	
Year Moved In	1960	
Year of Entry	1890	1940–1960

Exhibit 3

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

Simple Demographics Often Identify People Uniquely

Latanya Sweeney
Carnegie Mellon University
latanya@andrew.cmu.edu

This work was funded in part by H. John Heinz III School of Public Policy and Management at Carnegie Mellon University and by a grant from the U.S. Bureau of Census.

Copyright © 2000 by Latanya Sweeney. All rights reserved.

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

1. Abstract

In this document, I report on experiments I conducted using 1990 U.S. Census summary data to determine how many individuals within geographically situated populations had combinations of demographic values that occurred infrequently. It was found that combinations of few characteristics often combine in populations to uniquely or nearly uniquely identify some individuals. Clearly, data released containing such information about these individuals should not be considered anonymous. Yet, health and other person-specific data are publicly available in this form. Here are some surprising results using only three fields of information, even though typical data releases contain many more fields. It was found that 87% (216 million of 248 million) of the population in the United States had reported characteristics that likely made them unique based only on {5-digit ZIP, gender, date of birth}. About half of the U.S. population (132 million of 248 million or 53%) are likely to be uniquely identified by only {place, gender, date of birth}, where place is basically the city, town, or municipality in which the person resides. And even at the county level, {county, gender, date of birth} are likely to uniquely identify 18% of the U.S. population. In general, few characteristics are needed to uniquely identify a person.

2. Introduction

Data holders often collect person-specific data and then release derivatives of collected data on a public or semi-public basis after removing all explicit identifiers, such as name, address and phone number. Evidence is provided in this document that this practice of de-identifying data and of ad hoc generalization are not sufficient to render data anonymous because combinations of attributes often combine uniquely to re-identify individuals.

2.1. Linking to re-identify de-identified data

In this subsection, I will demonstrate how linking can be used to re-identify de-identified data. The National Association of Health Data Organizations (NAHDO) reported that 44 states have legislative mandates to collect hospital level data and that 17 states have started collecting ambulatory care data from hospitals, physicians offices, clinics, and so forth [1]. These data collections often include the patient's *ZIP* code, *birth date*, *gender*, and *ethnicity* but no explicit identifiers like *name* or *address*. The leftmost circle in Figure 1 contains some of the data elements collected and shared.

For twenty dollars I purchased the voter registration list for Cambridge Massachusetts and received the information on two diskettes [2]. The rightmost circle in Figure 1 shows that these data included the name, address, ZIP code, birth date, and gender of each voter. This information can be linked using *ZIP*, *birth date* and *gender* to the medical information, thereby linking diagnosis, procedures, and medications to particularly named individuals. The question that remains of course is how unique would such linking be.

In general I can say that the greater the number and detail of attributes reported about an entity, the more likely that those attributes combine uniquely to identify the entity. For example, in the voter list, there were 2 possible values for gender and 5 possible five-digit ZIP codes; birth dates were within a range of 365 days for 100 years. This gives 365,000 unique values, but there were only 54,805 voters.

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

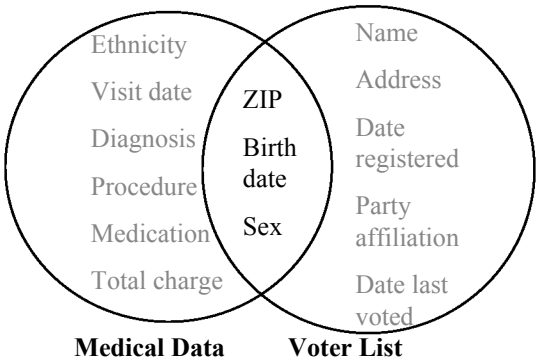


Figure 1 Linking to re-identify data

2.2. Publicly and semi-publicly available health data

As mentioned in the previous subsection, most states (44 of 50 or 88%) collect hospital discharge data [3]. Many of these states have subsequently distributed copies of these data to researchers, sold copies to industry and made versions publicly available. While there are many possible sources of patient-specific data, these represent a class of data collections that are often publicly and semi-publicly available.

#	Field description	Size
1	HOSPITAL ID NUMBER	12
2	PATIENT DATE OF BIRTH (MMDDYYYY)	8
3	SEX	1
4	ADMIT DATE (MMDYYYY)	8
5	DISCHARGE DATE (MMDDYYYY)	8
6	ADMIT SOURCE	1
7	ADMIT TYPE	1
8	LENGTH OF STAY (DAYS)	4
9	PATIENT STATUS	2
10	PRINCIPAL DIAGNOSIS CODE	6
11	SECONDARY DIAGNOSIS CODE - 1	6
12	SECONDARY DIAGNOSIS CODE - 2	6
13	SECONDARY DIAGNOSIS CODE - 3	6
14	SECONDARY DIAGNOSIS CODE - 4	6
15	SECONDARY DIAGNOSIS CODE - 5	6
16	SECONDARY DIAGNOSIS CODE - 6	6
17	SECONDARY DIAGNOSIS CODE - 7	6
18	SECONDARY DIAGNOSIS CODE - 8	6
19	PRINCIPAL PROCEDURE CODE	7
20	SECONDARY PROCEDURE CODE - 1	7
21	SECONDARY PROCEDURE CODE - 2	7
22	SECONDARY PROCEDURE CODE - 3	7
23	SECONDARY PROCEDURE CODE - 4	7
24	SECONDARY PROCEDURE CODE - 5	7
25	DRG CODE	3
26	MDC CODE	2
27	TOTAL CHARGES	9
28	ROOM AND BOARD CHARGES	9
29	ANCILLARY CHARGES	9
30	ANESTHESIOLOGY CHARGES	9
31	PHARMACY CHARGES	9
32	RADIOLOGY CHARGES	9
33	CLINICAL LAB CHARGES	9
34	LABOR-DELIVERY CHARGES	9
35	OPERATING ROOM CHARGES	9
36	ONCOLOGY CHARGES	9
37	OTHER CHARGES	9
38	NEWBORN INDICATOR	1
39	PAYER ID 1	9
40	TYPE CODE 1	1
41	PAYER ID 2	9
42	TYPE CODE 2	1
43	PAYER ID 3	9
44	TYPE CODE 3	1
45	PATIENT ZIP CODE	5
46	Patient Origin COUNTY	3
47	Patient Origin PLANNING AREA	3
48	Patient Origin HSA	2
49	PATIENT CONTROL NUMBER	
50	HOSPITAL HSA	2

Figure 2 IHCCCC Research Health Data

The Illinois Health Care Cost Containment Council (IHCCCC) is the organization in the State of Illinois that collects and disseminates health care cost data on hospital visits in Illinois. IHCCCC reports more than 97% compliance by Illinois hospitals in providing the information

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

[4]. Figure 2 contains a sample of the kinds of fields of information that are not only collected, but also disseminated.

Of the states mentioned in the NAHDO report, 22 of these states contribute to a national database called the State Inpatient Database (*SID*) sponsored by the Agency for Healthcare Research and Quality (AHRQ). A copy of each patient's hospital visit in these states is sent to AHRQ for inclusion in *SID*. Some of the fields provided in *SID* are listed in Figure 3 along with the compliance of the 13 states that contributed to *SID*'s 1997 data [5].

Field	Comments	#states	%states
Patient Age	years	13	100%
Patient Date of birth	month, year	5	38%
Patient Gender		13	100%
Patient Racial background		11	85%
Patient ZIP	5-digit	9	69%
Patient ID	encrypted (or scrambled)	3	23%
Admission date	month, year	8	62%
Admission day of week		12	92%
Admission source	emergency, court/law, etc	13	100%
Birth weight	for newborns	5	38%
Discharge date	month, year	7	54%
Length of stay		13	100%
Discharge status	routine, death, nursing home, etc	13	100%
Diagnosis Codes	ICD9, from 10 to30	13	100%
Procedure Codes	from 6 to 21	13	100%
Hospital ID	AHA#	12	92%
Hospital county		12	92%
Primary payer	Medicare, insurance, self-pay, etc	13	100%
Charges	from 1 to 63 categories	11	85%

Figure 3 Some data elements for AHRQ's State Inpatient Database (13 participating states)

State	Month and Year of Birth date	Age
Arizona	Yes	Yes
California		Yes
Colorado		Yes
Florida		Yes
Iowa	Yes	Yes
Massachusetts		Yes
Maryland		Yes
New Jersey		Yes
New York	Yes	Yes
Oregon	Yes	Yes
South Carolina		Yes
Washington		Yes
Wisconsin	Yes	Yes

Figure 4 Age information provided by states to SID

Figure 4 lists the states reported in Figure 3 that provide the *month and year of birth* and the *age* for each patient.

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

The remainder of this document provides experimental results from summary data that show how demographics often combine to make individuals unique or almost unique in data like these.

2.3. A single attribute

The frequency with which a single characteristic occurs in a population can help identify individuals based on unusual or outlying information. Consider a frequency distribution of birth years found in the list of registered voters. It is not surprising to see fewer people present with earlier birth years. Clearly, a person born in 1900 is unusual and by implication less anonymous in data.

2.4. More than one attribute

What may be more surprising is that combinations of characteristics can combine to occur even less frequently than the characteristics appear alone.

ZIP	Birth	Gender	Race
60602	7/15/54	m	Caucasian
60140	2/18/49	f	Black
62052	3/12/50	f	Asian

Figure 5 Data that looks anonymous

Consider Figure 5. If the three records shown were part of a large and diverse database of information about Illinois residents, then it may appear reasonable to assume that these three records would be anonymous. However, the 1990 federal census [6] reports that the ZIP (postal code) 60602 consisted primarily of a retirement community in the Near West Side of Chicago and therefore, there were very few people (less than 12) of an age under 65 living there. The ZIP code 60140 is the postal code for Hampshire, Illinois in Dekalb county and reportedly there were only two black women who resided in that town. Likewise, 62052 had only four Asian families. In each of these cases, the uniqueness of the combinations of characteristics found could help re-identify these individuals.

Race	Birth	Gender	ZIP	Problem
Black	09/20/65	m	02141	short of breath
Black	02/14/65	m	02141	chest pain
Black	10/23/65	f	02138	hypertension
Black	08/24/65	f	02138	hypertension
Black	11/07/64	f	02138	obesity
Black	12/01/64	f	02138	chest pain
White	10/23/64	m	02138	chest pain
White	03/15/65	f	02139	hypertension
White	08/13/64	m	02139	obesity
White	05/05/64	m	02139	short of breath
White	02/13/67	m	02138	chest pain
White	03/21/67	m	02138	chest pain

Figure 6 De-identified data

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

As another example, Figure 6 contains de-identified data. Each row contains information about a distinct person, so information about 12 people is reported. The table contains the following fields of information {*Race/Ethnicity*, *Date of Birth*, *Gender*, *ZIP*, *Medical Problem*}.

In Figure 6, there is information about an equal number of African Americans (listed as *Black*) as there are Caucasian Americans (listed as *White*) and an equal number of men (listed as *m*) as there are women (listed as *f*), but in combination, there appears only one Caucasian female.

2.5. Learned from the examples

These examples demonstrate that in general, the frequency distributions of combinations of characteristics have to be examined in combination with respect to the entire population in order to determine unusual values and cannot be generally predicted from the distributions of the characteristics individually. Of course, obvious predictions can be made from extreme distributions --such as values that do not appear in the data will not appear in combination either.

3. Background of definitions and terms

> **Definition (informal). Person-specific data** Collections of information whose granularity of details are specific to an individual are termed *person-specific data*. More generally, in *entity-specific data*, the granularity of details is specific to an entity.

Example. Person-specific data

Figure 5 and Figure 6 provide examples of person-specific data. Each row of these tables contains information related to one person.

The idea of anonymous data is a simple one. The term "anonymous" means that the data cannot be linked or manipulated to confidently identify the individual who is the subject of the data.

> **Definition (informal). Anonymous data** *Anonymous data* implies that the data cannot be manipulated or linked to confidently identify the entity that is the subject of the data.

Most people understand that there exist explicit identifiers, such as name and address, which can provide a direct means to communicate with the person. I term these *explicit identifiers*; see the informal definition below.

> **Definition (informal). Explicit identifier** An *explicit identifier* is a set of data elements, such as {*name*, *address*} or {*name*, *phone number*}, for which there exists a direct communication method, such as email, telephone, postal mail, etc., where with no additional information, the designated person could be directly and uniquely contacted.

A common incorrect belief is that removing all explicit identifiers such as name, address and phone number from the data renders the result anonymous. I refer to this instead as *de-identified data*; see the informal definition below.

Definition (informal). De-identified data *De-identified data* result when all explicit identifiers, such as name, address, or phone number are removed, generalized or replaced with a made-up alternative.

Example. De-identified data

Figure 5 and Figure 6 provide examples of de-identified person-specific data. There are no explicit identifiers in these data.

Because a combination of characteristics can combine uniquely for an individual, it can provide a means of recognizing a person and therefore serve as an identifier. In the literature, such combinations were nominally introduced as *quasi-identifiers* [7] and *identificates* [3-58] with no supporting evidence provided as to how identifying specific combinations might be. Extending beyond the literature and its casual use in the literature, I term such a combination a quasi-identifier and informally define it below. I then examine specific quasi-identifiers found within publicly and semi-publicly available data and compute their general ability to uniquely associate with particular persons in the U.S. population.

Definition (informal). Quasi-identifier A quasi-identifier is a set of data elements in entity-specific data that in combination associates uniquely or almost uniquely to an entity and therefore can serve as a means of directly or indirectly recognizing the specific entity that is the subject of the data.

Example. Quasi-identifier

A quasi-identifier whose values are unique for all the records in Figure 6 is {*ZIP*, *gender*, *Birth*}.

In the next section, I will show that {*ZIP*, *gender*, *Birth*} is a unique quasi-identifier for most people in the U.S. population.

The term *table* is really quite simple and is synonymous with the casual use of the term data collection. It refers to data that are conceptually organized as a 2-dimensional array of rows (or records) and columns (or fields). A database is considered to be a set of one or more tables.

Definition (informal). Table, tuple and attribute A *table* conceptually organizes data as a 2-dimensional array of rows (or records) and columns (or fields). Each row (or record) is termed a *tuple*. A tuple contains a relationship among the set of values associated with an entity. Tuples within a table are not necessarily unique. Each column (also known as a field or data element) is called an *attribute* and denotes a field or semantic category of information that is a set of possible values; therefore, an attribute is also a domain. Attributes within a table are unique. So by observing a table, each row is an ordered n-tuple of values $<d_1, d_2, \ldots, d_n>$ such that each value d_j is in the domain of the j-th column, for j=1, 2, …, n where n is the number of columns.

In mathematical set theory, a relation corresponds with this tabular presentation; the only difference is the absence of column names. Ullman provides a detailed discussion of relational database concepts [9].

Examples of tables

Figure 5 provides an example of a person-specific table with attributes {*ZIP*, *Birth*, *Gender*, *Race*}. Each tuple concerns information about a single person. Figure 6 provides an example of a person-specific table with attributes {*Race*, *Birth*, *Gender*, *ZIP*, *Problem*}.

Unfortunately, the terminology with respect to data collections is not the same across communities and diverse communities have an interest in this work. In order to accommodate these different vocabularies, I provide the following thesaurus of interchangeable terms. In general, *data collection*, *data set* and *table* refer to the same representation of information though a data collection may have more than one table. The terms *record*, *row* and *tuple* all refer to same kind of information. Finally, the terms *data element*, *field*, *column* and *attribute* refer to the same kind of information. For brevity, from this point forward, I will use the more formal database terms of table, tuple and attribute. I do allow the tuples of a table to appear in a "sorted" order on occasion and such cases pose a slight deviation from its more formal meaning. These uses are explicitly noted.

4. Methods

4.1. Census Tables

Information from the 1990 US Census made available on the Web [10] and on CDROM [11] and from the U.S. Postal Service [12] was loaded into Microsoft Access and the following tables produced and used with Microsoft Excel.

1. ZIP census table provides 1990 federal census information summarized by each ZIP (postal code) in the United States.

2. Place census table provides 1990 federal census information summarized by place name (town, city, municipality, or postal facility name).

3. County census table provides 1990 federal census information summarized by US counties.

Figure 7 contains a list of attributes (or data elements) for each of these tables. The name and description of each attribute is listed and a "yes" appears in the column that associates the attribute to the ZIP, Place or County table in which the attribute appears. Information for all 50 states and the District of Columbia were provided. For example, values associated with the attribute *Tot_pop* in the ZIP table are the total numbers of individuals reported as living in each corresponding *ZIP*. Each tuple (or row) in the table corresponds to a unique ZIP.

Given a particular geographical specification such as ZIP, place or county, the number of people reported as residing in the noted geographical area is reported by age subdivision in the ZIP, Place and County tables. The age subdivisions are: under 12 years of age (denoted as *Aunder12*), between 12 and 18 years of age (denoted as *A12to18*), between 19 and 24 years of age (denoted as *A19to24*), between 25 and 34 years of age (denoted as *A25to34*), between 35 and 44 years of age (denoted as *A35to44*), between 45 and 54 years of age (denoted as *A45to54*),

between 55 and 64 years of age (denoted as *A55to64*) or more than 65 years of age (denoted as *A65Plus*).

Field	Description	ZIP	PLACE	COUNTY
StateID	State Code	yes	yes	yes
ZIP	5-digit ZIP	yes	**NO**	**NO**
Place	Name of Incorporated Place	**NO**	yes	**NO**
CoName	County Name	**NO**	**NO**	yes
Tot_Pop	Total Population	yes	yes	yes
AUnder12	Population Under Age 12 Years	yes	yes	yes
A12to18	Population Age 12-18 Years	yes	yes	yes
A19to24	Population Age 19-24 Years	yes	yes	yes
A25to34	Population Age 25-34 Years	yes	yes	yes
A35to44	Population Age 35-44 Years	yes	yes	yes
A45to54	Population Age 45-54 Years	yes	yes	yes
A55to64	Population Age 55-64 Years	yes	yes	yes
A65Plus	Population Age 65 Years and up	yes	yes	yes

Figure 7 1990 Census attributes in ZIP, Place, County tables

4.2. ZIPNameGIS Table

ZIP information provided from the U.S. Postal Service included place, which is a name of a town, city, municipality or postal facility uniquely assigned to a ZIP code. This information was loaded directly to provide the ZIPNameGIS table. The attributes (or data elements) for the ZIPNameGIS table are {*StateID*, *ZIP*, *State*, *POName*, *longitude*, *latitude*, *population*}.

The Place table was constructed by linking the ZIP table to the ZIPNameGIS table on *ZIP*. Results were then grouped by *POName* (respecting state designations) so that population information from multiple ZIP codes were grouped together by the city or town in which the ZIP code referred. Finally, the Place table was generated by collapsing these groupings into single entries that contained the sum of the population values reported for all ZIP codes corresponding to the same place.

During the process, 3 ZIP codes were found to cross state lines and therefore, be listed in two states. To avoid this duplication, the following assignments were made: (1) ZIP code 32530 refers to Pinetta in both Florida and Georgia. The Georgia entry was removed from Place; (2) ZIP code 42223 refers to Fort Campbell in both Kentucky and Tennessee. The Tennessee entry was removed from Place; and, (3) ZIP code 63673 refers to Saint Mary in both Illinois and Missouri. The Missouri entry was removed from Place.

4.2.1. Schemas of shared data

Figure 2 and Figure 3 contain descriptions of publicly and semi-publicly available hospital discharge data. Below are some quasi-identifiers found in those data that also appear in the census data. The experiments reported in this document estimate the uniqueness of values associated with these quasi-identifiers given the occurrences reported in the census data.

1. Illinois Research Health Data.

The Illinois Research Health Data (R_{rod}) is described in Figure 2. Among the attributes listed there, I consider QI_{rod} = {*date of birth*, *gender*, *5-digit ZIP*} to be a quasi-identifier within R_{rod}.

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

2. AHRQ's State Inpatient Database

The Agency for Healthcare Research and Quality's State Inpatient Database (R_{SID}) is described in part in Figure 3. Among the attributes listed there, I consider QI_{SID1} = {*month and year of birth*, *gender*, *5-digit ZIP*} to be a quasi-identifier within data released by some states and I consider QI_{SID2} = {*age*, *gender*, *5-digit ZIP*} to be a quasi-identifier within data released by other states.

4.3. Design and procedures

The experiments reported in the next section can be generally described in terms of values attributes can assume. Let $T(A_1,\ldots,A_n)$ be an entity-specific table and let Q_T be a quasi-identifier of T. Q_T is represented as a finite set of attributes $\{A_i,\ldots,A_j\} \subseteq \{A_1,\ldots,A_n\}$. I write $|A_m|$ to represent the finite number of values A_m can assume. So, the number of distinct possible values that be assigned to Q_T, written $|Q_T|$, is: $|Q_T| = |A_i| * |A_{i+1}| * \ldots * |A_j|$.

Example.
Given Q_{dob}={*date of birth*, *gender*}, then $|Q_{dob}| = 365 * 76 * 2 = 55{,}480$ because there are 365 days in a year, an expected lifetime of 76 years, and 2 genders.

In this document, I am concerned with a person-specific table $T(A_1,\ldots,Z,\ldots,A_n)$ that includes a *geographic attribute Z*. Values assigned to a *geographic attribute* are specific to the residences of people. Examples of geographic attributes include 5-digit ZIP codes, names of cities and towns, and names of counties in which people reside. Let *U* be the universe of all people and the person-specific table $Geo[z_i, A_r, \ldots, A_s)$ contain all or almost all of the people of *U* having $Z=z_i$. I say Geo_{zi} is a *population register* for z_i. And, $T[A_1,\ldots Z_i,\ldots A_n]$ is a pseudo-random sample drawn from $Geo[z_i, A_r, \ldots, A_s]$. Unique and unusual combinations of characteristics found in Geo with respect to z_i can be no less unique or unusual when recorded in T. Therefore, the probability distribution of combinations of characteristics found in Geo limits the values those combinations of characteristics can assume in T. Determining unique and unusual combinations of characteristics within a residential domain is a counting problem.

Theorem. Generalized Dirichlet drawer principle [13]
(also known as the Generalized pigeonhole principle)

If *N* objects are distributed in *k* boxes, then there is at least one box containing at least $\lceil N / k \rceil$ objects.

Proof.
Suppose that none of the boxes contain more than $\lceil N / k \rceil$ -1 objects. Then, the total number of objects is at most: $k * (\lceil N / k \rceil - 1) < k * (((N / k) + 1) - 1) = N$
This has the inequality $\lceil N / k \rceil < (N / k) + 1$
This is a contradiction because there are a total of *N* objects.

Example.
Given a random sample of 500 people, there are at least $\lceil 500 / 365 \rceil = 2$ people with the same birthday because there are 365 possible birthdays.

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

Let z_i be a 5-digit ZIP code. I write *population*(z_i) to denote the number of people who reside in z_i and *population*(z_i) ≡ |Geo_{zi}|. If *population*(Z_i) > |Q_{dob}|, then by the generalized pigeonhole principle, a tuple $t \in \mathsf{R}_{rod}$[*date of birth*, *gender*, z_i] would not uniquely correspond to one person. In these cases, I say t[A_1, …, *date of birth*, *gender*, z_i, ..., An] is not likely to be uniquely identifiable. On the other hand, if *population*(z_i) ≤ |Q_{dob}| then by the generalized pigeonhole principle, a tuple $t \in \mathsf{R}_{rod}$[*date of birth*, *gender*, z_i] would likely relate to only one person. In these cases, I say t[A_1, …, *date of birth*, *gender*, z_i, ..., An] is likely to be uniquely identifiable. This is the general approach to the experiments reported in the next section though each differs in terms of attribute specification.

4.3.1. Subdivision analyses

The analyses of the identifiability of geographically situated populations are based on age-based divisions within a geographic attribute. Let age subdivision a be either *Aunder12*, *A12to18*, *A19to24*, *A25to34*, *A35to44*, *A45to54*, *A55to64*, or *A65Plus*. The quasi-identifier Q_a has the same attributes as Q_{dob} but values which *date of birth* can assume are limited by a. That is, |Q_a| is the number of possible distinct values that can be assigned to Q_a. I say |Q_a| is the *threshold* for Q_{dob} with respect to age subdivision a.

Example.
Given Q_{dob}={*date of birth*, *gender*} and age subdivision a = *A19to24*, then |Q_a| = 365 * 2 * 6 = 4380 because there are 365 birthdays, 2 genders and 6 years between the ages of 19 to 24, inclusive.

Number of subjects uniquely identified in a subdivision of a geographical area (ID_{aZi})

Given a value for a geographic attribute, written z_i, and an age subdivision a, I write *population*(z_i, a) as the number of people residing in z_i with an age within a. The number of people considered uniquely identified by a and Z_i, written ID_{aZi}, is determined by the rule:

if *population*(z_i, a) ≥ |Q_a|, then ID_{aZi}= *population*(z_i, a)
else ID_{aZi} = 0.

By extension, the percentage of people residing in z_i considered uniquely identified (written ID_{zi}) with respect to the set of age subdivisions is computed as:

$$ID_{Zi} = \frac{population(z_i) - \sum_{a=AUnder12}^{A65Plus} ID_{aZi}}{population(z_i)}$$

4.3.2. Statistics on geographical areas

Statistics are reported on geographic regions. Given a geographic attribute Z, let $Region_Z$ = {z_i | z_i ∈ Z } and *AgeDivs* = {*Aunder12*, *A12to18*, *A19to24*, *A25to34*, *A35to44*, *A45to54*, *A55to64*, *A65Plus*}. That is, $Region_Z$ is a set of values that can be assigned to the geographic

attribute Z and *AgeDivs* is a set of age subdivisions. $Region_Z$ is partitioned into NotIDSet and IDSet based on age subdivision a∈*AgeDivs* such that:

$$\text{NotIDSet}_{Za} = \{(z_i,a) \mid z_i \in Region_Z \text{ and } population(z_i, a) > |Q_a|\ \}$$
$$\text{IDSet}_{Za} = \{(z_i,a) \mid z_i \in Region_Z \text{ and } population(z_i, a) \le |Q_a|\ \}$$

The population of NotIDSet_Z is not considered uniquely identifiable by values of Q_{dob}. The population of IDSet_Z is considered uniquely identifiable by values of Q_{dob}. In the experiments, the following statistics are reported.

Maximum subpopulation(NotIDSet_{Za}) = max(*population*(z_1, *a*), …, *population*(z_y, *a*)), where $(z_i,a) \in \text{NotIDSet}_{Za}$

Maximum subpopulation(IDSet_{Za}) = max(*population*(z_1, *a*), …, *population*(z_y, *a*)), where $(z_i,a) \in \text{IDSet}_{Za}$

Minimum subpopulation(NotIDSet_{Za}) = min(*population*(z_1, *a*), …, *population*(z_y, *a*)), where $(z_i,a) \in \text{NotIDSet}_{Za}$

Minimum subpopulation(IDSet_{Za}) = min(*population*(z_1, *a*), …, *population*(z_y, *a*)), where $(z_i,a) \in \text{IDSet}_{Za}$

$$Average\ \ subpopulation(\text{NotIDSet}_{Za}) = \frac{\sum_{(z_i,a)\in NotIDSet} population(z_i,a)}{|\text{NotIDSet}_{Za}|}$$

$$Average\ \ subpopulation(\text{IDSet}_{Za}) = \frac{\sum_{(z_i,a)\in IDSet} population(z_i,a)}{|\text{IDSet}_{Za}|}$$

$$Number\ of\ geographical\ areas(\text{NotIDSet}_{Za}) = |\text{NOTIDSet}_{Za}|$$

$$Number\ of\ geographical\ areas(\text{IDSet}_{Za}) = |\text{IDSet}_{Za}|$$

$$Percentage\ of\ geographical\ areas(\text{NotIDSet}_{Za}) = \frac{|\text{NotIDSet}_{Za}|}{|\text{NotIDSet}_{Za}| + |\text{IDSet}_{Za}|}$$

$$Percentage\ of\ geographical\ areas(\text{IDSet}_{Za}) = \frac{|\text{IDSet}_{Za}|}{|\text{NotIDSet}_{Za}| + |\text{IDSet}_{Za}|}$$

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

State		AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
AL	4,040,587	699,554	425,425	369,639	652,466	585,299	422,565	363,033	522,606
AK	544,698	123,789	53,662	46,478	111,790	101,699	55,887	29,236	22,157
AZ	3,665,228	678,439	352,557	333,055	639,702	530,192	354,711	299,372	477,200
AR	2,350,725	410,665	246,486	197,424	361,268	328,397	244,096	212,573	349,816
CA	29,755,274	5,436,303	2,722,076	2,904,739	5,738,645	4,645,553	2,955,455	2,231,171	3,121,332
CO	3,293,771	599,278	305,595	282,268	617,333	570,797	340,276	249,924	328,300
CT	3,287,116	517,724	275,158	295,271	588,185	509,760	360,488	294,866	445,664
DE	666,168	113,963	58,980	64,726	119,782	100,110	68,367	59,570	80,670
DC	606,900	80,760	45,404	71,605	122,777	94,984	62,648	51,050	77,672
FL	12,686,788	1,931,088	1,041,486	1,010,156	2,102,614	1,778,994	1,283,728	1,235,820	2,302,902
GA	6,478,847	1,171,969	659,386	623,625	1,182,367	1,014,579	678,987	495,259	652,675
HI	1,108,229	195,278	98,594	104,537	203,466	178,406	109,493	93,778	124,677
ID	1,006,749	207,979	115,708	81,770	154,087	149,338	98,910	77,819	121,138
IL	11,429,942	2,012,780	1,102,499	1,021,458	2,003,217	1,702,509	1,179,345	974,035	1,434,099
IN	5,543,954	975,582	568,654	510,374	919,924	819,577	572,585	481,329	695,929
IA	2,776,442	487,879	271,630	240,359	430,947	397,287	272,959	249,594	425,787
KS	2,474,885	457,755	236,911	216,092	416,003	363,571	234,451	208,146	341,956
KY	3,673,969	626,236	383,356	337,585	610,721	549,204	380,791	320,712	465,364
LA	4,219,973	836,481	458,677	387,821	710,773	606,119	412,186	340,483	467,433
ME	1,226,626	210,082	117,015	104,754	205,713	194,139	123,745	108,198	162,980
MD	4,771,143	812,147	409,957	431,840	901,956	774,414	528,246	395,946	516,637
MA	6,011,978	933,306	506,033	613,116	1,104,645	914,852	605,951	514,398	819,677
MI	9,295,222	1,671,777	930,841	850,016	1,583,364	1,408,199	950,316	793,711	1,106,998
MN	4,370,288	815,963	409,705	377,084	783,562	666,480	428,315	343,315	545,864
MS	2,573,216	495,074	298,599	240,546	403,754	351,197	249,684	213,117	321,245

Figure 8 Population by state and age group, part 1

State		AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
MO	5,113,266	897,590	490,067	436,468	855,640	734,252	524,756	457,095	717,398
MT	799,065	150,406	83,457	57,351	123,913	128,067	81,522	67,930	106,419
NE	1,577,600	294,659	156,790	130,613	259,709	229,478	148,720	134,711	222,920
NV	1,201,833	208,695	100,891	102,609	223,599	192,324	138,893	107,621	127,201
NH	1,109,252	195,970	98,977	100,411	205,815	183,649	111,387	88,059	124,984
NJ	7,730,188	1,217,936	681,960	664,059	1,366,267	1,200,167	850,983	718,589	1,030,227
NM	1,515,069	307,898	160,598	123,983	259,975	229,577	149,712	120,808	162,518
NY	17,990,026	2,891,618	1,615,696	1,664,461	3,148,965	2,720,452	1,944,539	1,642,487	2,361,808
NC	6,628,637	1,074,691	637,603	662,849	1,152,229	1,008,277	705,099	585,832	802,057
ND	637,713	119,767	65,036	57,151	104,833	90,808	56,215	53,132	90,771
OH	10,846,581	1,899,661	1,064,732	957,750	1,805,063	1,619,291	1,115,355	978,701	1,406,028
OK	3,145,585	563,941	318,809	267,411	514,663	452,308	326,770	278,089	423,594
OR	2,842,321	495,834	265,630	225,488	455,371	476,343	297,101	235,423	391,131
PA	11,881,643	1,892,957	1,074,128	1,041,626	1,918,168	1,739,212	1,224,867	1,160,974	1,829,711
RI	1,003,211	155,439	86,271	102,680	174,149	146,571	97,958	89,156	150,987
SC	3,486,703	616,373	363,140	339,600	596,534	526,103	357,747	291,077	396,129
SD	695,133	137,110	71,070	56,976	109,919	96,063	61,962	59,623	102,410
TN	4,896,046	812,832	484,155	452,701	823,042	740,485	530,654	433,773	618,404
TX	16,984,748	3,320,887	1,776,426	1,578,004	3,118,515	2,548,657	1,649,538	1,284,825	1,707,896
UT	1,722,850	430,959	226,933	167,637	275,853	224,715	139,656	107,405	149,692
VT	562,758	99,365	53,099	53,049	95,880	92,804	57,274	45,118	66,169
VA	6,184,493	1,030,088	564,690	616,835	1,147,609	991,563	670,457	500,955	662,296
WA	4,866,692	878,141	444,693	417,468	861,441	804,413	504,238	380,725	575,573
WV	1,792,969	279,885	192,881	148,808	262,961	270,784	191,957	176,960	268,733
WI	4,891,452	887,426	472,270	437,743	825,056	726,753	478,819	412,492	650,893
WY	453,588	92,123	49,716	33,980	75,462	74,182	45,541	35,539	47,045
USA	248,418,140	43,454,102	23,694,112	22,614,049	43,429,692	37,582,954	25,435,905	21,083,554	31,123,772

Figure 9 Population by state and age group, part 2

Different experiments have different age and geographic attributes. See Figure 11 for a list of all 13 experiments identified as A through M. So, Q_{dob} and Z_i, as used above, are representative of several quasi-identifiers that have varying specifications. In experiment B through experiment E, $Z_i \in$ {ZIP codes in USA in which people reside}. In experiment F through experiment I, $Z_i \in$ {Cities, municipalities, towns and recognized post office names in the USA}. Finally, in experiment J through experiment M, $Z_i \in$ {Counties in the USA}. Similarly, in experiments B, F, and J, Q_{dob} = {*date of birth*, *gender*}. In experiments C, G and K, Q_{dob} = {*month and year of birth*, *gender*}. In experiments D, H and L, Q_{dob} = {*year of birth*, *gender*}. Finally, in experiments E, I and M, Q_{dob} = {*2 year age subdivision*, *gender*}.

For completeness, Figure 8 and Figure 9 report the total population per state of each age group. These values are used to compute percentages throughout this document unless otherwise noted.

4.4. Special data elements

This section compares age and year of birth values, as well as, 5-digit ZIP codes, places and counties.

4.4.1. Age versus Year of Birth

Values for an *age* attribute do not necessarily translate to known values for a *year of birth* attribute. There are two cases to consider. If there exists a date to which values for *age* can be referenced, then corresponding values for *year of birth* can be confidently computed. For example, in SID, states calculate the patient's age in years at the time of admission [14]. Because both the computed *age* and the *date of admission* are released, the patient's year of birth can be confidently determined. In experiment D, H and L, I examine *age* as providing a distinct year of birth, and so QI_{SID2} = {*age*, *gender*, *5-digit ZIP*} can be considered as QI_{SID2} = {*year of birth*, *gender*, *5-digit ZIP*}.

On the other hand, if values for *date of admission* were not released, values for *age* would be calendar year specific. In such cases, data are collected with respect to a particular calendar year (that is known) but not a particular day within that year. As a result, each value for *age* corresponds to two possible values for each person's *year of birth*. During any given calendar year, a person reports two ages. The first age occurs before the person's birthday and the second occurs on and after the person's birthday. Because each person's birthday can appear at any time during the calendar year (in contrast to societies in which everyone's "birthday", in terms of determining age, occurs on the same day), two values can be inferred for *year of birth* from a recorded value for *age*. In the experiment E, I and M, I examine {*2 yr age subdivision*, *gender*, *5-digit ZIP*} in which the birth year is within a known 2-year range.

4.4.2. Comparison of 5-digit ZIP codes, Places and Counties

Figure 10 shows a comparison of 5-digit ZIP codes, places and counties in the United States. There are a total of 29,343 ZIP codes, 25,688 places and 3,141 counties. The state having the largest number of counties was Texas (with 254). The District of Columbia had the fewest number of counties (with 1). The average number of counties per state was 62 and the standard deviation was 47.

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

State	Number 5-digit ZIPs	Number Places	Number Counties	State	Number digit ZIPs	Number Places	Number Counties
AL	567	511	67	MO	993	899	115
AK	195	183	25	MT	315	309	57
AZ	270	178	15	NE	572	518	93
AR	578	563	75	NV	104	66	17
CA	1,515	1,071	58	NH	218	212	10
CO	414	330	63	NJ	540	490	21
CT	263	224	8	NM	276	258	33
DE	53	46	3	NY	1,594	1,369	62
DC	24	2	1	NC	705	624	100
FL	804	463	67	ND	387	384	53
GA	636	561	159	OH	1,007	854	88
HI	80	70	5	OK	586	511	77
ID	244	233	44	OR	384	344	36
IL	1,236	1,147	102	PA	1,458	1,369	67
IN	675	597	92	RI	69	52	5
IA	922	889	99	SC	350	313	46
KS	713	646	105	SD	383	377	66
KY	810	772	120	TN	583	505	95
LA	469	408	64	TX	1,672	1,234	254
ME	410	408	16	UT	205	181	29
MD	419	378	24	VT	243	243	14
MA	473	404	14	VA	820	729	136
MI	875	768	83	WA	484	397	39
MN	877	809	87	WV	655	646	55
MS	363	342	82	WI	714	666	72
				WY	141	135	23
				USA	29,343	25,688	3,141
				max	1,672	1,369	254
				min	24	2	1
				avg	575	504	62
				stdev	401	337	47

Figure 10 Number of 5-digit ZIP codes, Places and Counties by State

5. Results

In the previous sections, I defined terminology and introduced the materials that will be used. In this section, I report on experiments I conducted to estimate the number of unique occurrences for various combinations of demographic attributes that are typically released in publicly and semi-publicly available data.

Experiment A: Uniqueness of {*ZIP*, *gender*, *date of birth*} assume uniform age distribution
Experiment B: Uniqueness of {*ZIP*, *gender*, *date of birth*} based on actual age distribution
Experiment C: Uniqueness of {*ZIP*, *gender*, *month and year of birth*}
Experiment D: Uniqueness of {*ZIP*, *gender*, *age*}
Experiment E: Uniqueness of {*ZIP*, *gender*, *2yr age range*}
Experiment F: Uniqueness of {*place/city*, *gender*, *date of birth*}
Experiment G: Uniqueness of {*place/city*, *gender*, *month and year of birth*}
Experiment H: Uniqueness of {*place/city*, *gender*, *age*}
Experiment I: Uniqueness of {*place/city*, *gender*, *2yr age range*}
Experiment J: Uniqueness of {*county*, *gender*, *date of birth*}
Experiment K: Uniqueness of {*county*, *gender*, *month and year of birth*}
Experiment L: Uniqueness of {*county*, *gender*, *age*}
Experiment M: Uniqueness of {*county*, *gender*, *2yr age range*}

Figure 11 List of 13 experiments

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

A total of 13 experiments were conducted [15]. These are identified below. Only experiment B, C, D, F and J are briefly reported in this document. Figure 32 contains a summary of results from all 13 experiments.

5.1. Experiment B: Uniqueness of {*ZIP*, *gender*, *date of birth*}

Recall, Illinois Research Health Data named ROD provides an example of shared data that contains demographic attributes; in particular, QI_{rod} = {*date of birth*, *gender*, *5-digit ZIP*}. This experiment shows that medical conditions included in these data can be attributed uniquely to one person in most cases.

5.1.1. Experiment B Design

Step 1. Use ZIP table for each of the 50 states and the District of Columbia. Step 2. Figure 12 contains the thresholds for *Q*={*gender*, *date of birth*} specific to each age subdivision. Step 3. Report statistical measurements computed from the table in step 1 using the thresholds determined in step 2. Figure 13 and Figure 14 report the results.

Q = {*gender*, *date of birth*}	
$\|Q_{AUnder12}\|$	= 2 * 365 * 12 = 8,760
$\|Q_{A12to18}\|$	= 2 * 365 * 7 = 5,110
$\|Q_{A19to24}\|$	= 2 * 365 * 6 = 4,380
$\|Q_{A25to34}\|$	= 2 * 365 * 10 = 7,300
$\|Q_{A35to44}\|$	= 2 * 365 * 10 = 7,300
$\|Q_{A45to54}\|$	= 2 * 365 * 10 = 7,300
$\|Q_{A55to64}\|$	= 2 * 365 * 10 = 7,300
$\|Q_{A65Plus}\|$	= 2 * 365 * 12 = 8,760

Figure 12 Number of possible values for each age subdivision {*gender*, *date of birth*}

5.1.2. Experiment B Results

Figure 13 and Figure 14 show the results from applying the 3 steps of experiment B to each state, the District of Columbia and the entire United States. The percentages computed for each locale appear in the column named "RANGE %ID_pop." The last row in Figure 14 reports the results of applying the 3 steps of experiment B to all ZIP codes in the United States. As shown, 87.1% of the population of the United States is likely to be uniquely identified by values of {*gender*, *date of birth*, *ZIP*} when age subdivisions are considered.

During the analysis of experiment B, many interesting ZIP codes were found. Here are a few. The ZIP code 11794 in the State of New York is small and extremely homogenous. 4666 of its total population of 5418 (or 86%) are in the age subdivision of 19 to 24. This is the home of the State University of New York at Sony Brook. The ZIP code 10475 in the State of New York reportedly has a larger population of 37077, but people are distributed somewhat evenly across the age subdivisions making the population in each range less than its corresponding threshold. The ZIP code 01701 in the Commonwealth of Massachusetts reportedly has a population of 65,001, which is the largest population for a ZIP code in the state. In experiment A, any person residing in that ZIP code would NOT have been considered likely to be uniquely identified by {*gender*, *date of birth*, *ZIP*}; however, only the subpopulation between the ages of 19 and 44 in

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

that ZIP code is large enough not to be considered uniquely identified by {*gender*, *date of birth*, *ZIP*}. Persons residing in that ZIP code, who are not in that age subdivision, are less common and considered likely to be uniquely identified by {*gender*, *date of birth*, *ZIP*} even though the population in the entire ZIP code is the largest in the state.

State	#ZIPs	Population	RANGE %population	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
AL	567	4,040,587	99%	100.0%	100.0%	89.7%	98.7%	100.0%	100.0%	100.0%	100.0%
AK	195	544,698	100%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
AZ	270	3,665,228	82%	82.3%	90.1%	67.4%	64.3%	88.8%	100.0%	100.0%	80.7%
AR	578	2,350,725	98%	97.8%	100.0%	87.1%	95.3%	100.0%	100.0%	100.0%	100.0%
CA	1,515	29,755,274	71%	62.4%	73.1%	54.9%	47.2%	70.0%	96.8%	99.6%	96.8%
CO	414	3,293,771	92%	89.7%	96.2%	85.0%	81.1%	92.1%	100.0%	100.0%	100.0%
CT	263	3,287,116	91%	94.3%	98.1%	76.1%	76.2%	88.9%	100.0%	100.0%	97.8%
DE	53	666,168	91%	100.0%	100.0%	72.0%	66.7%	100.0%	100.0%	100.0%	100.0%
DC	24	606,900	64%	62.0%	74.9%	32.5%	47.6%	55.3%	100.0%	84.9%	85.1%
FL	804	12,686,788	91%	93.9%	95.8%	87.5%	85.2%	94.3%	98.6%	99.2%	83.6%
GA	636	6,478,847	90%	90.4%	93.5%	80.4%	77.8%	87.6%	100.0%	100.0%	100.0%
HI	80	1,108,229	74%	62.5%	94.4%	56.7%	55.9%	71.9%	100.0%	100.0%	83.7%
ID	244	1,006,749	99%	100.0%	100.0%	85.6%	100.0%	100.0%	100.0%	100.0%	100.0%
IL	1,236	11,429,942	75%	73.0%	76.4%	59.2%	60.1%	73.9%	90.3%	93.9%	86.7%
IN	675	5,543,954	94%	94.3%	95.2%	80.4%	85.4%	94.7%	100.0%	100.0%	100.0%
IA	922	2,776,442	98%	100.0%	100.0%	78.9%	98.0%	100.0%	100.0%	100.0%	100.0%
KS	713	2,474,885	98%	100.0%	100.0%	83.1%	94.1%	100.0%	100.0%	100.0%	100.0%
KY	810	3,673,969	98%	100.0%	100.0%	85.7%	97.5%	98.6%	100.0%	100.0%	100.0%
LA	469	4,219,973	91%	89.8%	91.7%	80.4%	83.6%	93.0%	100.0%	100.0%	100.0%
ME	410	1,226,626	98%	100.0%	100.0%	86.3%	96.3%	100.0%	100.0%	100.0%	100.0%
MD	419	4,771,143	83%	84.8%	94.1%	79.2%	63.7%	80.2%	93.8%	100.0%	88.7%
MA	473	6,011,978	91%	95.7%	97.9%	73.5%	74.8%	92.8%	100.0%	100.0%	98.8%
MI	875	9,295,222	85%	80.5%	84.7%	72.5%	74.5%	83.2%	98.2%	99.1%	98.3%
MN	877	4,370,288	95%	96.2%	100.0%	81.8%	87.7%	97.4%	100.0%	100.0%	100.0%
MS	363	2,573,216	98%	98.2%	98.1%	88.3%	100.0%	97.8%	100.0%	100.0%	100.0%

Figure 13 Uniqueness of {*ZIP*, *Gender*, *Date of birth*} respecting age distribution, part 1

State	#ZIPs	Population	RANGE %population	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
MO	993	5,113,266	94%	94.4%	98.8%	86.9%	86.8%	92.1%	100.0%	100.0%	97.3%
MT	315	799,065	98%	100.0%	100.0%	78.9%	100.0%	100.0%	100.0%	100.0%	100.0%
NE	572	1,577,600	99%	100.0%	100.0%	90.2%	100.0%	100.0%	100.0%	100.0%	100.0%
NV	104	1,201,833	86%	79.5%	94.3%	79.5%	66.9%	88.3%	94.6%	100.0%	100.0%
NH	218	1,109,252	97%	100.0%	100.0%	94.1%	88.5%	100.0%	100.0%	100.0%	100.0%
NJ	540	7,730,188	92%	92.6%	93.1%	88.0%	79.8%	92.9%	99.1%	100.0%	94.1%
NM	276	1,515,069	88%	86.1%	89.0%	88.6%	71.6%	82.4%	100.0%	100.0%	100.0%
NY	1,594	17,990,026	76%	74.3%	77.3%	64.1%	60.0%	72.1%	88.3%	93.4%	85.5%
NC	705	6,628,637	94%	98.1%	96.4%	77.5%	86.4%	96.5%	98.8%	100.0%	100.0%
ND	387	637,713	96%	100.0%	100.0%	68.5%	91.9%	100.0%	100.0%	100.0%	100.0%
OH	1,007	10,846,581	92%	92.2%	94.7%	82.4%	82.5%	93.6%	100.0%	100.0%	98.5%
OK	586	3,145,585	97%	96.7%	100.0%	85.2%	93.5%	96.7%	100.0%	100.0%	100.0%
OR	384	2,842,321	97%	100.0%	100.0%	89.5%	90.6%	93.1%	100.0%	100.0%	100.0%
PA	1,458	11,881,643	91%	90.5%	94.0%	80.1%	82.2%	90.3%	99.3%	99.4%	94.3%
RI	69	1,003,211	92%	94.4%	100.0%	71.1%	84.2%	94.9%	100.0%	100.0%	94.2%
SC	350	3,486,703	91%	90.0%	95.1%	74.8%	79.5%	95.0%	97.9%	100.0%	100.0%
SD	383	695,133	96%	92.7%	100.0%	81.4%	91.6%	100.0%	100.0%	100.0%	100.0%
TN	583	4,896,046	93%	93.7%	94.8%	80.5%	87.1%	93.5%	100.0%	100.0%	100.0%
TX	1,672	16,984,748	88%	85.0%	89.1%	78.8%	76.5%	90.0%	100.0%	100.0%	100.0%
UT	205	1,722,850	87%	75.8%	80.0%	78.0%	90.2%	92.6%	100.0%	100.0%	100.0%
VT	243	562,758	98%	100.0%	100.0%	80.1%	100.0%	100.0%	100.0%	100.0%	100.0%
VA	820	6,184,493	87%	88.2%	91.6%	71.9%	75.5%	82.7%	97.8%	100.0%	100.0%
WA	484	4,866,692	92%	94.6%	100.0%	82.8%	82.5%	87.2%	100.0%	100.0%	100.0%
WV	655	1,792,969	97%	96.7%	96.4%	90.2%	95.7%	96.4%	100.0%	100.0%	96.5%
WI	714	4,891,452	92%	88.9%	97.7%	77.6%	86.4%	92.6%	100.0%	100.0%	100.0%
WY	141	453,588	98%	100.0%	100.0%	79.2%	100.0%	100.0%	100.0%	100.0%	100.0%
USA	29,343	248,418,140	87%	85.8%	90.2%	75.0%	75.1%	87.0%	97.8%	99.0%	95.3%

Figure 14 Uniqueness of {*ZIP*, *Gender*, *Date of birth*} respecting age distribution, part 2

Figure 15 plots the percentage of the population considered identifiable in each ZIP code in the United States based on experiment B's criteria. The horizontal axis represents the

population that resides in the ZIP code. The vertical axis represents the percentage of the population considered uniquely identified by values of *Q*={*date of birth*, *gender*, *5-digit ZIP*} for a particular ZIP code. The criteria for computing the percentage of the population considered identifiable in experiment B is based on binary decisions, where each decision considers whether a sufficient number of people in a particular age subdivision reside in a particular ZIP code. If so, that sub-population is not considered identifiable; otherwise, its entire sub-population is considered identifiable.

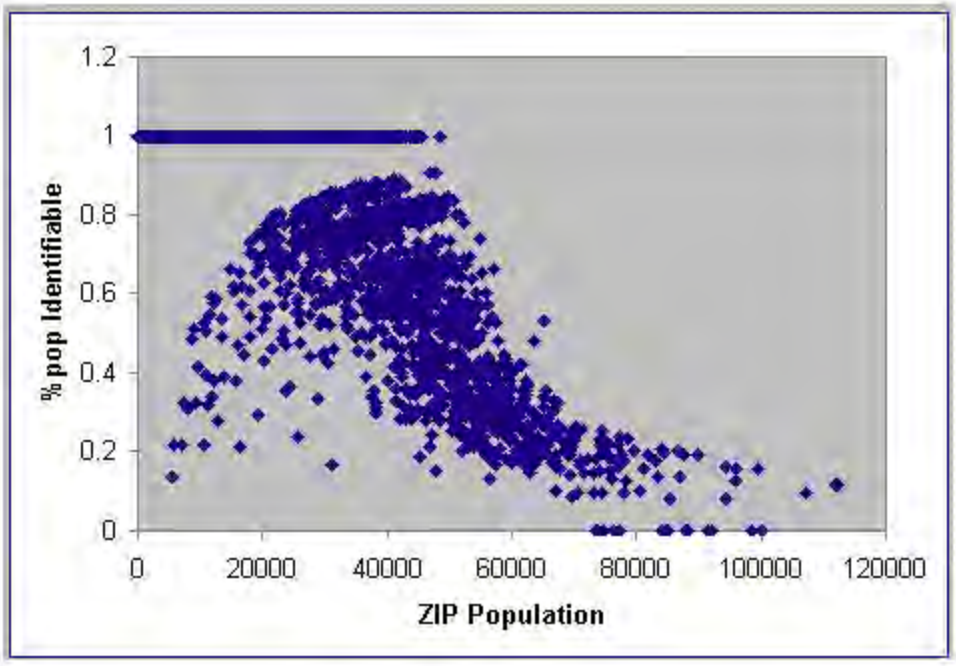


Figure 15 Percentage of Population Identifiable Based on Age subdivisions in ZIP Population

Most ZIP codes (27697 of 29212 or 95%) in the United States that have people listed as residing within them do not have enough people in any age subdivision to consider any such sub-population as identifiable. This is evidenced in Figure 15 by the appearance of dots where the *%pop identifiable* is 1. The largest population having *%pop identifiable* = 1 consists of 48,549 total people. There are very few ZIP codes (15 of 29212) in Figure 15 having sufficient numbers of people in each age subdivision that each such sub-population is not considered uniquely identifiable. This is evidenced in Figure 15 by the appearance of dots where the *%pop identifiable* is 0. The largest population having *%pop identifiable* = 0 has 99,995 people and the smallest has 73,321.

The ZIP code having the largest population, ZIP 60623 with 112,167 people, has a percentage of its population considered identifiable in Figure 15 as being only 11%. It is not 0% because there are insufficient numbers of people above the age of 55 living there despite the large number of people residing in the ZIP code. The point representing this ZIP code in Figure 15 is the rightmost point shown.

The lowest leftmost point shown in Figure 15 corresponds to ZIP 11794, which was discussed earlier. It has a total population of 5418 people and consists primarily of people between the ages of 19 and 24 (4666 of 5418 or 86%). Despite having a small population, the people residing there are very homogenous in terms of age and so the percentage of its population

considered identifiable based on experiment B's criteria is only 13%. It is clear from these examples that population size alone is not an absolute predictor of the identifiability of the people residing within. Care must be taken to model the population as precisely as possible to insure privacy protection.

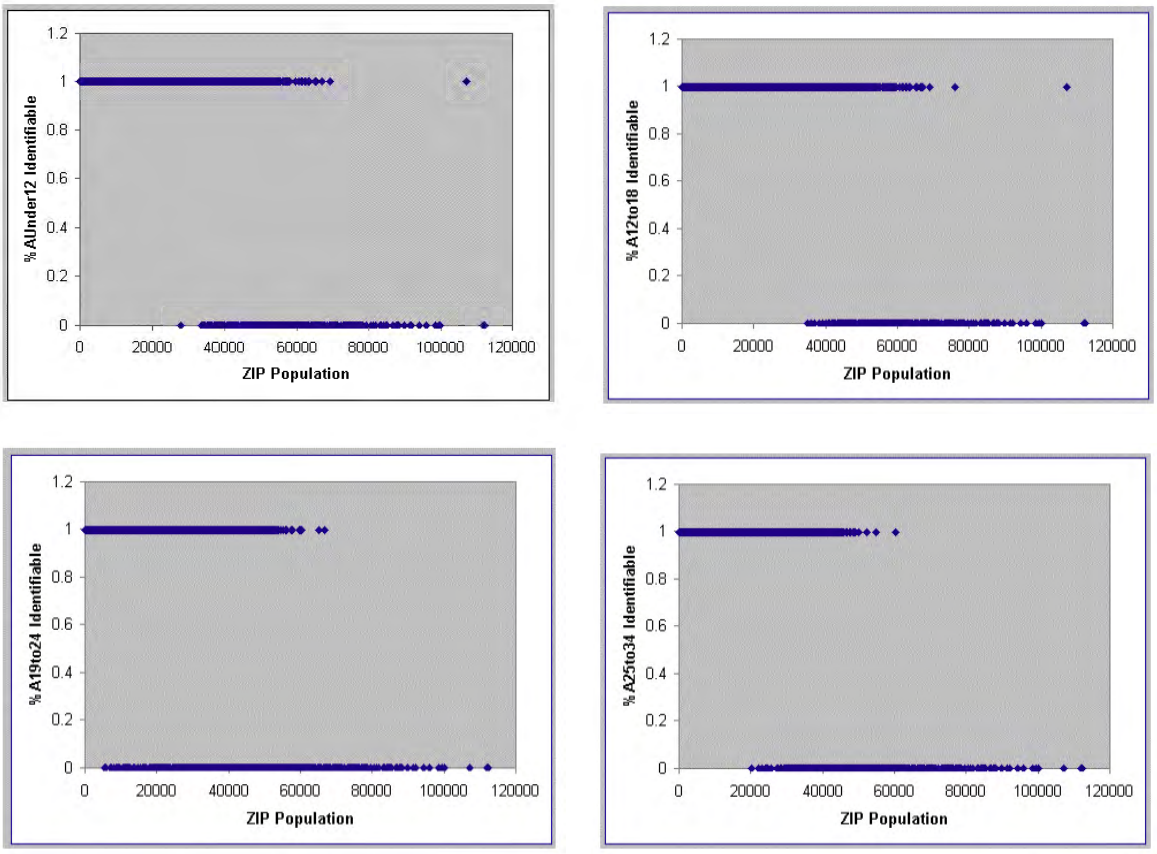


Figure 16 Percentage of Age-based Populations Identifiable within ZIP Population, Part 1

Recall the computation of the percentage of the population considered uniquely identified by values of *Q*={*date of birth*, *gender*, *5-digit ZIP*} for a particular ZIP code in experiment B is based on a composite of binary decisions. Each binary decision concerns the number of people residing within a specific ZIP code in a particular age subdivision. Figure 16 and Figure 17 show plots of the percentage of sub-populations considered identifiable in each ZIP code in the United States based on experiment B's criteria. The horizontal axis represents the population that resides in the ZIP code. The vertical axis represents the percentage of the population considered uniquely identified by values of *Q*={*date of birth*, *gender*, *5-digit ZIP*} for a particular ZIP code and a particular age subdivision. If a sufficient number of people within an age subdivision are reported as residing in a particular ZIP code, then that sub-population is considered identifiable; otherwise, the entire sub-population is not considered identifiable.

Figure 18 provides statistical highlights from the plots in Figure 16 and Figure 17. The topmost table provides statistics on ZIP codes in which the number of people within the noted age subdivision is less than or equal to the threshold for that subdivision. In these cases, the sub-population within the ZIP code is considered uniquely identifiable; that is, *%pop_Identifiable* = 1 for that age subdivision and ZIP code. The bottom table provides statistics in cases where *%pop_Identifiable* < 1. In these ZIP codes, the number of people within the noted age subdivision

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

is greater than the threshold for that subdivision; therefore, this subdivision is not considered uniquely identifiable.

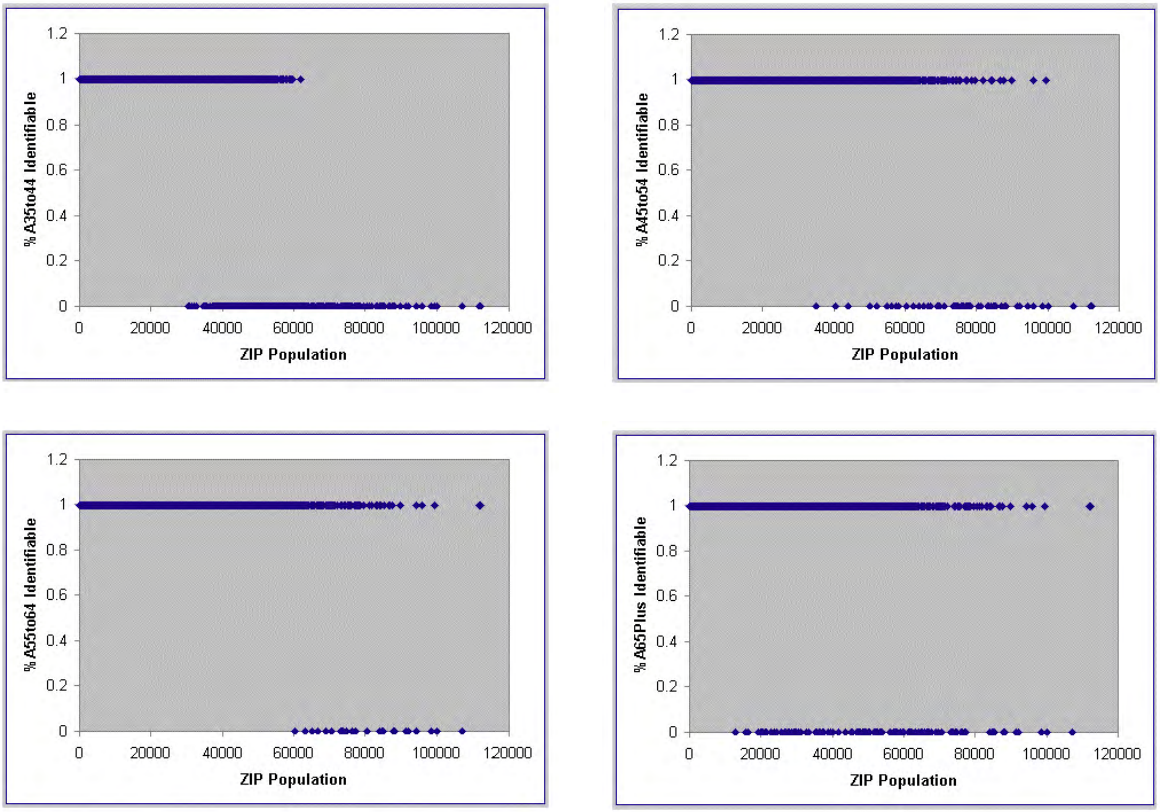


Figure 17 Percentage of Age-based Populations Identifiable within ZIP Population, Part 2

Sub-population considered uniquely identifiable (<= *threshold, IDSet*)

	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
Max ZIP population	107197	107197	66722	60388	62031	99420	112167	112167
Min ZIP population	1	1	1	1	1	1	1	1
Average ZIP population	7615	7873	7332	6911	7596	8358	8442	8311
standard deviation	19452	10915	10070	9227	10393	11938	12165	11956
Number of ZIP codes	28675	28860	28352	28105	28665	29148	29187	29081
Percentage ZIP codes	98.2%	98.8%	97.1%	96.2%	98.1%	99.8%	99.9%	99.6%

Sub-population NOT considered uniquely identifiable (> *threshold, NotIDSet*)

	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
Max ZIP population	112167	112167	112167	112167	112167	112167	107197	107197
Min ZIP population	28294	35092	5418	20211	30577	34860	60388	12890
Average ZIP population	55958	60254	47153	48944	56072	74798	80513	51313
standard deviation	12770	13036	17178	12681	13157	15961	12304	20367
Number of ZIP codes	537	352	860	1107	547	64	25	131
Percentage ZIP codes	1.8%	1.2%	2.9%	3.8%	1.9%	0.2%	0.1%	0.4%

Figure 18 Statistical highlights from Figure 16 and Figure 17

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

5.2. Experiment C: Uniqueness of {*ZIP*, *gender*, *month and year of birth*}

This experiment (referred to as experiment C) is motivated by the Agency for Healthcare Research and Quality's State Inpatient Database (R_{SID}), which is described in part in Figure 3. Among the attributes listed there, I consider QI_{SID1} = {*month and year of birth*, *gender*, *5-digit ZIP*} to be a quasi-identifier within data released by some states. This experiment attempts to characterize the identifiability of QI_{SID1}.

5.2.1. Experiment C Design

Step 1. Use ZIP table for each of the 50 states and the District of Columbia. Step 2. Figure 19 contains the thresholds for *Q*={*gender*, *month and year of birth*} specific to each age subdivision. Step 3. Report statistical measurements computed from the table in step 1 using the thresholds determined in step 2. Figure 20 and Figure 21 report the results.

Q3 = {*gender*, *month and year of birth*}		
$\|Q3_{AUnder12}\|$	= 2 * 12 * 12	= 288
$\|Q3_{A12to18}\|$	= 2 * 12 * 7	= 168
$\|Q3_{A19to24}\|$	= 2 * 12 * 6	= 144
$\|Q3_{A25to34}\|$	= 2 * 12 * 10	= 240
$\|Q3_{A35to44}\|$	= 2 * 12 * 10	= 240
$\|Q3_{A45to54}\|$	= 2 * 12 * 10	= 240
$\|Q3_{A55to64}\|$	= 2 * 12 * 10	= 240
$\|Q3_{A65Plus}\|$	= 2 * 12 * 12	= 288

Figure 19 Number of possible values for each age subdivision for {*gender*, *month and year of birth*}

5.2.2. Experiment C Results

Figure 20 and Figure 21 show the results of applying the 3 steps of experiment C to each state, the District of Columbia (as just reported) and the entire United States. The percentage of people residing in each locale likely to be uniquely identifiable based on {*gender*, *month and year of birth*, *ZIP*} appear in the column named "MonYr %ID_pop." For example, 18.1% of the population of Iowa (see Figure 20) and 26.5% of the population of North Dakota (see Figure 21) are likely to be uniquely identifiable based on {*gender*, *month and year of birth*, *ZIP*}.

The next to last row in Figure 21 labeled "USA" reports the results of applying the 3 steps of experiment C to all ZIP codes in the United States. As shown, 3.7% of the population of the United States is likely to be uniquely identified by values of {*gender*, *month and year of birth*, *ZIP*}. The last row in Figure 21 labeled "%ID_pop" displays the percentage of people in each age subdivision who are likely to be uniquely identified by values of {*gender*, *month and year of birth*, *ZIP*}. For example, it reports that 5% of the population of persons residing in the United States between the ages of 45 and 54 are likely to be uniquely identifiable based on {*gender*, *month and year of birth*, *ZIP*}.

Figure 22 plots the percentage of the population considered identifiable in each ZIP code in the United States based on experiment C's criteria. The horizontal axis represents the population that resides in the ZIP code. The vertical axis represents the percentage of the

population considered uniquely identified by values of QI_{SID1} = {*month and year of birth*, *gender*, *5-digit ZIP*} for a particular ZIP code. This is the same as the approach used in experiment B.

State	MonYr %ID_pop	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
AL	3.8%	22,253	11,325	10,982	18,197	19,285	22,443	24,806	24,254
AK	10.7%	12,416	6,542	4,826	9,045	7,633	6,253	5,522	6,063
AZ	1.4%	6,804	3,888	4,386	6,786	5,968	7,091	7,095	8,120
AR	11.4%	41,221	23,185	20,274	34,340	35,164	35,248	35,440	42,675
CA	0.8%	33,588	19,440	16,982	27,467	26,335	31,331	33,500	34,743
CO	3.7%	18,174	10,214	8,764	14,721	14,523	16,946	17,965	21,333
CT	1.2%	5,203	2,845	3,097	4,102	3,675	5,104	7,135	7,514
DE	0.9%	867	557	257	653	652	960	715	1,627
DC	0.2%	275	72	26	180	95	66	57	404
FL	0.6%	10,862	6,777	6,548	8,311	9,208	11,647	11,760	13,330
GA	2.7%	19,935	11,272	11,318	18,321	22,193	26,345	31,161	34,905
HI	1.6%	1,767	1,242	1,602	1,911	1,795	2,797	3,645	3,469
ID	8.9%	11,922	7,146	6,950	11,657	11,988	12,404	12,220	15,587
IL	4.4%	75,604	42,727	40,364	62,012	63,393	68,919	70,997	77,971
IN	4.0%	28,592	16,297	17,739	25,328	25,849	33,632	34,730	36,884
IA	18.1%	82,724	44,905	34,644	70,040	64,634	65,878	65,808	72,916
KS	12.1%	46,345	25,207	20,797	36,178	38,319	40,822	41,630	49,544
KY	8.3%	48,404	24,728	23,501	37,727	39,465	41,358	43,680	46,346
LA	2.8%	15,800	8,567	8,553	13,180	13,922	17,090	18,399	22,675
ME	15.5%	29,727	16,098	14,462	23,099	23,470	26,896	26,041	30,713
MD	2.1%	14,087	7,843	8,086	11,105	11,093	13,739	16,099	20,297
MA	1.1%	8,446	5,949	5,540	6,291	6,191	10,006	12,702	12,847
MI	2.4%	27,008	16,914	18,153	22,223	25,106	33,248	37,570	40,591
MN	9.0%	59,128	34,860	28,225	49,369	52,048	54,780	53,583	60,926
MS	4.4%	12,939	7,915	8,487	12,557	14,378	17,937	18,845	20,676

Figure 20 Uniqueness of {*ZIP*, *Gender*, *Month and year of birth*} respecting age distribution, part 1

Of the ZIP codes reported in Figure 22, about half (13,871 of 29,212 or 47%) have sufficient numbers of people in each age subdivision so that values of QI_{SID1} = {*month and year of birth*, *gender*, *5-digit ZIP*} are not likely to be uniquely identifying; in these cases, *%pop identifiable* = 0. Values of QI_{SID1} for about one third (9103 of 29212 or 31%) of the ZIP codes are considered uniquely identifying in all age subdivisions; in these cases, *%pop identifiable* = 1. The remaining ZIP codes (6238 of 29212 or 21%) have sub-populations in which values of QI_{SID1} are uniquely identifiable for some age subdivisions but not for others.

Figure 23 provides statistical highlights from the plot in Figure 22. The topmost table provides statistics on ZIP codes in which the number of people within the noted age subdivision is less than or equal to the threshold for that subdivision. In these cases, the sub-population within the ZIP code is considered uniquely identifiable; that is, *%pop_Identifiable* = 1 for that age subdivision and ZIP code. The bottom table provides statistics in cases where *%pop_Identifiable* < 1. In these ZIP codes, the number of people within the noted age subdivision is greater than the threshold for that subdivision; therefore, this subdivision is not considered uniquely identifiable. The method for computing these statistics was described earlier in the Methods section (on page 11).

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

State	MonYr %ID_pop	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
MO	8.2%	65,966	37,847	31,629	52,566	53,596	57,098	56,566	65,194
MT	15.5%	18,771	11,741	7,717	16,581	16,326	16,280	16,432	19,924
NE	18.2%	46,646	27,556	17,763	38,678	40,574	34,699	37,697	43,232
NV	2.0%	4,320	2,035	1,983	3,341	2,977	2,516	2,705	4,256
NH	7.5%	11,934	7,545	6,001	8,773	7,859	12,067	13,156	15,851
NJ	0.6%	6,760	4,693	3,510	3,811	4,642	5,846	8,238	8,142
NM	5.2%	11,169	6,307	5,208	10,048	10,235	10,844	11,340	14,141
NY	2.3%	54,792	33,243	31,443	45,160	49,560	61,882	68,223	76,979
NC	2.4%	22,064	11,906	10,595	17,177	16,987	23,559	27,726	31,714
ND	26.5%	28,362	16,090	9,492	22,535	22,563	20,666	22,226	27,314
OH	2.2%	28,645	14,449	18,930	24,301	24,283	37,395	43,814	47,838
OK	7.1%	32,749	20,178	16,901	26,174	29,484	29,507	32,320	35,238
OR	4.2%	18,614	9,286	8,839	15,741	14,495	15,766	15,778	19,684
PA	3.5%	58,144	32,516	32,758	47,305	45,996	62,507	66,894	75,584
RI	0.9%	1,085	642	500	764	1,417	1,025	1,487	1,996
SC	2.3%	9,342	5,171	5,813	8,643	8,309	12,372	13,670	16,738
SD	25.9%	27,699	17,147	11,054	25,496	24,375	22,171	23,721	28,405
TN	3.4%	24,172	12,553	13,053	18,105	19,074	22,832	25,898	30,553
TX	2.3%	51,615	29,794	30,883	45,082	50,060	58,173	62,784	68,838
UT	3.4%	8,496	4,844	4,042	7,026	7,447	8,832	8,293	10,307
VT	21.9%	19,797	11,196	8,334	16,536	17,312	16,075	16,093	18,066
VA	4.4%	41,345	23,241	20,634	30,706	33,035	35,263	40,117	47,007
WA	2.6%	18,736	11,083	9,104	14,925	15,043	17,563	19,665	21,650
WV	15.5%	43,535	25,866	21,381	36,753	37,676	34,584	35,731	42,582
WI	5.4%	32,406	21,664	21,855	31,257	30,297	40,576	43,567	44,714
WY	10.1%	8,492	3,943	2,743	6,058	5,943	6,251	5,893	6,684
USA	**3.7%**	1,329,747	759,051	676,728	1,098,342	1,125,947	1,269,289	1,351,139	1,529,041
%ID_pop		3.1%	3.2%	3.0%	2.5%	3.0%	5.0%	6.4%	4.9%

Figure 21 Uniqueness of {*ZIP*, *Gender*, *Month and year of birth*} respecting age distribution, part 2

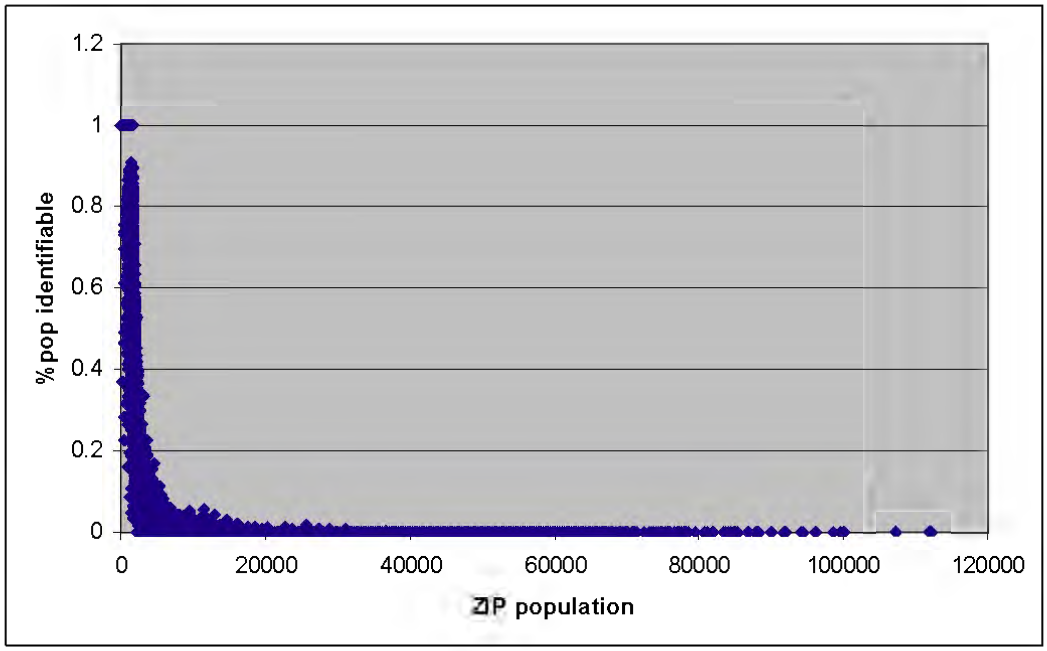


Figure 22 Percentage of Population Identifiable Based on Uniform Distribution of Ages in ZIP Population

The values reported as ZIP populations in Figure 22 are not the total number of people within the reported age subdivision residing in those ZIP codes but are just the numbers of people

residing in the ZIP code. For example, consider the values appearing in the "Aunder12" column in Figure 22. They report information about children under the age of 12 residing in 10,852 ZIP codes in the United States that had insufficient numbers of children to render corresponding values of QI_{SID1} = {*month and year of birth*, *gender*, *5-digit ZIP*} uniquely identifiable. Of these ZIP codes, the largest number of children of under the age of 12, residing in a ZIP code was 287. Some ZIP codes, who had people residing within them, had no children in this age. The average number of children in these ZIP codes was 123 with a standard deviation of 80.

Sub-population considered uniquely identifiable (<= *threshold, IDSet*)

	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
Max ZIP sub-population	287	167	143	239	239	239	239	287
Min ZIP sub-population	0	0	0	0	0	0	0	0
Average ZIP sub-population	123	71	53	101	102	96	95	118
standard deviation	80	47	40	66	66	66	66	80
Number of ZIP codes	10852	10725	12760	10883	11045	13202	14220	12905
Percentage ZIP codes	37.1%	36.7%	43.7%	37.3%	37.8%	45.2%	48.7%	44.2%

Sub-population NOT considered uniquely identifiable (> *threshold, NotIDSet*)

	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
Max ZIP sub-population	26914	15352	27123	24587	19543	15544	12205	25799
Min ZIP sub-population	288	168	144	240	240	240	240	288
Average ZIP sub-population	2294	1241	1333	2309	2007	1509	1316	1815
standard deviation	2530	1327	1690	2632	2096	1419	1174	1860
Number of ZIP codes	18360	18487	16452	18329	18167	16010	14992	16307
Percentage ZIP codes	62.9%	63.3%	56.3%	62.7%	62.2%	54.8%	51.3%	55.8%

Figure 23 Statistical highlights from Figure 20 and Figure 21

5.3. Experiment D: Uniqueness of {*ZIP*, *gender*, *age*}

In this experiment, I examine the identifiability of {*year of birth*, *gender*, *5-digit ZIP*} in the United States. Progressing through the results from the last three experiments, values referring to age became less specific and as expected, the values became less uniquely identifying. What may be surprising however is that these values remained uniquely identifying for some people.

The Agency for Healthcare Research and Quality's State Inpatient Database (SID; see Figure 3) motivated this experiment as well as experiment C. In addition to QI_{SID1} used in experiment C, SID also includes QI_{SID2} = {*age*, *gender*, *5-digit ZIP*} for some states in those data. Recall in section 4.4.1, I examine *age* as providing a distinct year of birth, and so QI_{SID2} = {*age*, *gender*, *5-digit ZIP*} can be considered as QI_{SID2} = {*year of birth*, *gender*, *5-digit ZIP*}.

5.3.1. Experiment D Design

Step 1. Use ZIP table for each of the 50 states and the District of Columbia. Step 2. Figure 24 contains the thresholds for Q={*gender*, *date of birth*} specific to each age subdivision. Step 3. Report statistical measurements computed from the table in step 1 using the thresholds determined in step 2. Figure 25 and Figure 26 report the results.

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

Q4 = {*gender*, *year of birth*}		
$\|Q4_{AUnder12}\|$	= 2 * 12	= 24
$\|Q4_{A12to18}\|$	= 2 * 7	= 14
$\|Q4_{A19to24}\|$	= 2 * 6	= 12
$\|Q4_{A25to34}\|$	= 2 * 10	= 20
$\|Q4_{A35to44}\|$	= 2 * 10	= 20
$\|Q4_{A45to54}\|$	= 2 * 10	= 20
$\|Q4_{A55to64}\|$	= 2 * 10	= 20
$\|Q4_{A65Plus}\|$	= 2 * 12	= 24

Figure 24 Number of possible values for each age subdivision for {*gender*, *year of birth*}

5.3.2. Experiment D Results

Figure 25 and Figure 26 show the results of applying the 3 steps of experiment D to each state, the District of Columbia (as just reported) and the entire United States. The percentage of people residing in each locale likely to be uniquely identifiable based on {*gender*, *year of birth*, *ZIP*} appears in the column named “BirthYr %ID_pop” and the number of people represented by the percentage appears in the column named "BirthYr #ID_pop". For example, 0.89% (or 5703 people) of the population of Iowa (see Figure 26) are likely to be uniquely identifiable by values of {*gender*, *year of birth*, *ZIP*}.

State	BirthYr %ID_pop	BirthYr Total	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
AL	0.02%	918	105	53	89	97	112	125	158	179
AK	0.70%	3,809	227	223	227	223	315	631	804	1,159
AZ	0.02%	638	68	31	23	53	98	98	96	171
AR	0.09%	2,121	452	138	264	208	248	312	349	150
CA	0.01%	4,229	541	319	362	461	336	540	678	992
CO	0.08%	2,752	287	224	346	336	201	447	426	485
CT	0.01%	474	69	55	36	52	30	108	63	61
DE	0.02%	158	18	13	21	28	36	5	10	27
DC	0.01%	46	6	-	-	-	-	-	16	24
FL	0.00%	512	76	63	9	5	43	90	121	105
GA	0.01%	780	83	29	91	101	56	120	182	118
HI	0.01%	165	28	11	9	33	42	12	20	10
ID	0.19%	1,943	259	148	205	255	258	310	248	260
IL	0.01%	1,401	167	111	148	141	123	246	255	210
IN	0.01%	746	82	27	54	88	84	89	131	191
IA	0.11%	3,106	278	305	647	182	249	583	535	327
KS	0.22%	5,482	575	446	924	571	594	1,017	750	605
KY	0.13%	4,722	671	309	280	528	448	697	966	823
LA	0.02%	870	118	48	75	118	84	135	169	123
ME	0.19%	2,296	293	217	190	287	228	280	331	470
MD	0.03%	1,275	152	119	96	156	179	187	194	192
MA	0.01%	499	83	50	51	35	25	58	100	97
MI	0.01%	920	124	133	134	151	71	133	120	54
MN	0.06%	2,709	365	214	439	421	265	326	335	344
MS	0.02%	462	54	23	21	39	26	57	136	106

Figure 25 Uniqueness of {*ZIP*, *Gender*, *Year of birth*} respecting age distribution, part 1

The next to last row in Figure 26 labeled "USA" reports the results of applying the 3 steps of experiment D to all ZIP codes in the United States. As shown, 0.04% (or 105,016 people) of the population of the United States is likely to be uniquely identified by values of {*gender*, *year of birth*, *ZIP*}. The last row in Figure 26 labeled "%ID_pop" displays the percentage of people in each age subdivision who are likely to be uniquely identified by values of {*gender*, *year of birth*, *ZIP*}. For example, it reports that 0.08% of the population of persons residing in the

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

United States between the ages of 55 and 64 are likely to be uniquely identified by values of {*gender*, *year of birth*, *ZIP*}.

State	BirthYr %ID_pop	BirthYr Total	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
MO	0.07%	3,403	451	320	402	312	371	549	531	467
MT	0.43%	3,465	399	263	405	433	362	534	492	577
NE	0.23%	3,560	241	241	717	325	387	676	455	518
NV	0.04%	439	77	35	39	47	47	62	57	75
NH	0.07%	777	154	62	106	56	81	111	100	107
NJ	0.01%	728	125	62	41	61	51	96	114	178
NM	0.22%	3,302	343	276	237	395	350	569	644	488
NY	0.03%	5,460	714	469	533	720	445	804	818	957
NC	0.02%	1,032	133	94	74	134	103	177	168	149
ND	0.89%	5,703	586	476	832	675	639	932	787	776
OH	0.00%	377	34	25	30	37	33	38	96	84
OK	0.06%	1,963	220	135	248	219	274	336	237	294
OR	0.07%	1,900	369	140	172	258	124	214	315	308
PA	0.03%	3,099	501	201	324	413	348	429	440	443
RI	0.01%	92	-	-	-	9	10	30	19	24
SC	0.01%	443	87	16	41	66	63	85	45	40
SD	0.63%	4,408	489	291	607	544	516	632	597	732
TN	0.02%	836	201	14	125	70	53	165	128	80
TX	0.03%	5,483	815	383	443	641	661	717	794	1,029
UT	0.08%	1,323	78	59	146	189	151	230	230	240
VT	0.20%	1,117	76	63	171	54	81	166	150	356
VA	0.06%	3,754	572	286	350	423	445	483	638	557
WA	0.03%	1,227	164	85	145	138	122	142	220	211
WV	0.30%	5,360	746	316	433	614	605	874	869	903
WI	0.02%	881	80	101	135	130	79	103	103	150
WY	0.41%	1,851	213	157	232	165	195	361	223	305
USA	**0.04%**	105,016	13,049	7,879	11,729	11,697	10,747	16,121	16,463	17,331
%ID_pop			0.03%	0.03%	0.05%	0.03%	0.03%	0.06%	0.08%	0.06%

Figure 26 Uniqueness of {*ZIP*, *Gender*, *Year of birth*} respecting age distribution, part 2

Most ZIP codes (25,705 of 29,212 or 88%) have sufficient numbers of people in each age subdivision so that values of QI_{SID2} = {*year of birth*, *gender*, *5-digit ZIP*} are not likely to be uniquely identifying; in these cases, *%pop identifiable* = 0. Values of QI_{SID2} for about one third (353 of 29212 or 1%) of the ZIP codes are considered uniquely identifying in all age subdivisions; in these cases, *%pop identifiable* = 1. The remaining ZIP codes (3154 of 29212 or 11%) have sub-populations in which values of QI_{SID2} are uniquely identifiable for some age subdivisions but not for all.

Figure 27 provides statistical highlights. The topmost table provides statistics on ZIP codes in which the number of people within the noted age subdivision is less than or equal to the threshold for that subdivision. In these cases, the sub-population within the ZIP code is considered uniquely identifiable; that is, *%pop_Identifiable* = 1 for that age subdivision and ZIP code. The bottom table provides statistics in cases where *%pop_Identifiable* < 1. In these ZIP codes, the number of people within the noted age subdivision is greater than the threshold for that subdivision; therefore, this subdivision is not considered uniquely identifiable.

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

Sub-population considered uniquely identifiable (<= *threshold, IDSet*)

	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
Max ZIP sub-population	24	14	12	20	20	20	20	24
Min ZIP sub-population	0	0	0	0	0	0	0	0
Average ZIP sub-population	11	6	5	10	9	10	9	11
standard deviation	8	5	4	7	7	7	7	8
Number of ZIP codes	1200	1342	2309	1210	1150	1651	1798	1584
Percentage ZIP codes	4.1%	4.6%	7.9%	4.1%	3.9%	5.7%	6.2%	5.4%

Sub-population NOT considered uniquely identifiable (> *threshold, NotIDSet*)

	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
Max ZIP sub-population	26914	15352	27123	24587	19543	15544	12205	25799
Min ZIP sub-population	25	15	13	21	21	21	21	25
Average ZIP sub-population	1551	850	840	1551	1339	922	768	1126
standard deviation	2291	1212	1460	2372	1914	1284	1057	1652
Number of ZIP codes	28012	27870	26903	28002	28062	27561	27414	27628
Percentage ZIP codes	95.9%	95.4%	92.1%	95.9%	96.1%	94.3%	93.8%	94.6%

Figure 27 Statistical highlights from Figure 25 and Figure 26

5.4. Experiment F: Uniqueness of {*place*/*city*, *gender*, *date of birth*}

This experiment examines the identifiability of {*date of birth*, *gender*, *place*}. While the number of places is expected to be less than the number of ZIP codes, the difference is not as dramatic as one would expect.

State	DOB %ID_pop	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
AL	74.31%	510,294	316,271	246,921	455,646	425,871	340,085	290,787	416,713
AK	67.62%	86,943	36,668	30,801	72,365	66,328	34,744	18,336	22,157
AZ	30.18%	207,821	117,371	79,857	154,789	150,173	121,318	116,660	158,254
AR	85.73%	355,634	221,013	144,471	278,355	286,099	217,119	197,637	314,833
CA	35.99%	1,705,016	1,032,675	785,915	1,266,384	1,411,260	1,494,618	1,350,466	1,663,710
CO	40.41%	221,248	124,459	100,826	189,908	196,192	164,375	145,456	188,554
CT	66.44%	355,973	208,871	144,966	296,959	320,087	299,897	249,481	307,714
DE	68.04%	78,966	40,675	32,116	63,018	69,766	49,625	52,013	67,054
DC	0.00%	-	-	26	-	-	-	-	-
FL	44.12%	866,146	523,124	416,970	743,419	783,719	697,379	690,365	875,685
GA	62.62%	737,096	425,884	331,861	601,348	569,614	501,763	393,910	495,444
HI	49.94%	89,975	69,406	41,139	80,566	82,216	68,636	55,413	66,056
ID	76.93%	147,599	93,691	50,482	116,729	113,067	83,114	66,524	103,285
IL	60.16%	1,205,138	698,921	490,199	976,815	965,017	842,088	731,266	966,748
IN	63.45%	610,004	362,468	272,124	485,926	499,979	431,504	366,413	488,978
IA	77.50%	375,417	218,025	141,276	310,173	302,724	238,696	219,669	345,691
KS	66.77%	295,043	167,547	111,512	236,104	229,189	182,750	160,132	270,086
KY	78.76%	513,045	319,232	234,139	451,331	419,197	325,073	277,950	369,257
LA	58.86%	474,999	271,968	196,903	380,395	336,651	278,656	233,811	310,514
ME	94.22%	201,167	117,015	82,913	184,342	184,857	123,745	108,198	153,502
MD	63.22%	542,516	299,174	256,363	432,696	456,506	379,792	307,456	341,639
MA	73.33%	738,432	409,915	351,483	610,144	673,586	526,058	440,426	658,804
MI	56.68%	912,385	535,570	393,345	760,515	737,677	656,494	551,937	720,202
MN	71.55%	582,951	327,576	213,712	462,644	439,233	358,955	299,529	442,243
MS	81.12%	386,515	232,392	164,750	307,447	278,994	231,718	197,189	288,516

Figure 28 Uniqueness of {*Place*, *Gender*, *Date of birth*} respecting age distribution, part 1

Step 1. Use ZIP table for each of the 50 states and the District of Columbia. Step 2. Figure 12 contains the thresholds for Q={*gender*, *date of birth*} specific to each age subdivision. Step 3. Report statistical measurements computed from the table in step 1 using the thresholds

determined in step 2. Figure 28 and Figure 29 report the results of applying the 3 steps of experiment F to each state, the District of Columbia and the entire United States.

The percentage of people residing in each locale likely to be uniquely identifiable by values of {*gender*, *date of birth*, *place*} appear in the column named "DOB %ID_pop." For example, 94.22% of the population of Maine (see Figure 28) and 74.99% of the population of Pennsylvania (see Figure 29) are likely to be uniquely identifiable by values of {*gender*, *date of birth*, *place*}. Vermont had the largest percentage of its population identifiable (98.12%). The District of Columbia had 0% identified. The state having the smallest percentage was Nevada with 26.48%. The average was 64.54% and the standard deviation was 17.88%.

State	DOB %ID_pop	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
MO	65.98%	575,534	345,340	253,443	490,825	454,370	395,626	347,346	509,243
MT	78.05%	111,323	63,624	30,390	86,536	94,856	73,526	67,930	95,497
NE	60.86%	173,370	100,557	63,607	137,330	136,238	98,945	92,145	157,885
NV	26.48%	48,890	29,379	17,274	44,040	48,251	49,077	36,910	44,428
NH	83.26%	164,556	84,043	75,108	158,945	156,196	94,268	79,579	110,913
NJ	75.46%	916,586	513,909	459,760	887,738	910,504	705,604	615,918	823,232
NM	58.82%	185,741	103,241	70,980	125,794	127,320	94,373	78,403	105,250
NY	50.89%	1,510,307	893,370	734,124	1,331,293	1,394,790	1,103,058	955,471	1,231,836
NC	66.99%	748,655	434,802	352,507	670,230	637,726	523,682	455,492	617,381
ND	89.24%	108,831	59,803	33,455	83,627	83,251	56,215	53,132	90,771
OH	65.65%	1,218,515	726,779	536,583	1,009,900	1,059,754	865,805	737,419	965,782
OK	64.24%	349,375	209,852	141,980	280,350	266,557	233,933	212,063	326,461
OR	64.29%	318,531	186,694	120,253	251,227	266,919	224,214	180,088	279,439
PA	74.99%	1,427,475	829,811	674,412	1,324,556	1,288,682	1,002,535	960,527	1,401,861
RI	55.57%	83,379	52,128	46,137	74,615	83,775	73,597	65,732	78,157
SC	67.65%	404,179	259,598	178,853	347,400	357,955	263,798	240,827	306,073
SD	81.02%	108,221	62,338	36,113	80,508	80,733	53,059	51,721	90,508
TN	64.98%	529,152	319,932	243,251	474,021	459,452	388,946	320,903	433,014
TX	44.27%	1,410,090	792,176	561,715	1,100,437	1,053,590	840,761	735,749	1,025,466
UT	56.43%	208,964	117,137	81,156	132,730	134,699	106,448	84,198	106,867
VT	98.12%	99,365	53,099	42,494	95,880	92,804	57,274	45,118	66,169
VA	58.50%	588,706	358,361	294,519	565,454	531,480	468,056	357,966	453,394
WA	53.56%	458,232	257,086	168,811	372,536	382,178	319,725	272,740	375,511
WV	90.95%	260,338	178,947	125,468	232,443	242,711	184,384	169,168	237,233
WI	68.27%	584,155	333,763	235,969	497,263	483,528	372,939	334,139	497,585
WY	79.05%	67,039	36,679	20,714	52,859	53,145	45,541	35,539	47,045
USA	**58.38%**	24,859,832	14,572,359	10,914,146	20,826,555	20,879,466	17,343,591	15,107,247	20,512,640
%ID_pop		57.2%	61.5%	48.3%	48.0%	55.6%	68.2%	71.7%	65.9%

Figure 29 Uniqueness of {*Place*, *Gender*, *Date of birth*} respecting age distribution, part 2

The next to last row in Figure 29 labeled "USA" reports the results of applying the 3 steps of experiment F to all places in the United States. As shown, 58.38% of the population of the United States is likely to be uniquely identified by values of {*gender*, *date of birth*, *place*}. The last row in Figure 29 labeled "%ID_pop" displays the percentage of people in each age subdivision who are likely to be uniquely identified by values of {*gender, date of birth*, *place*}. For example, it reports that 71.7% of the population of persons residing in the United States between the ages of 55 and 64 are likely to be uniquely identifiable based on {*gender*, *date of birth*, *place*}.

The place having the largest population was Chicago, Illinois, with 2,451,767 people. The place having the smallest population was Crooked Creek, Alaska that reports only one person of age 65 or more resides there. The average population for a place is 9,710 and the standard deviation is 44,149. There are a total of 25,585 places.

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

5.5. Experiment J: Uniqueness of {*county*, *gender*, *date of birth*}

This experiment examines the identifiability of {*date of birth*, *gender*, *county*}. Recall, there are a total of 29,343 ZIP codes, 25,688 places and 3,141 counties.

Step 1. Use ZIP table for each of the 50 states and the District of Columbia. Step 2. Figure 12 contains the thresholds for *Q*={*gender*, *date of birth*} specific to each age subdivision. Step 3. Report statistical measurements computed from the table in step 1 using the thresholds determined in step 2. Figure 30 and Figure 31 report the results of applying the 3 steps of experiment J to each state, the District of Columbia and the entire United States.

The percentage of people residing in each locale likely to be uniquely identifiable by values of {*gender*, *date of birth*, *county*} appear in the column named "DOB %ID_pop." For example, 58% of the population of Mississippi (see Figure 30) and 52% of the population of Nebraska (see Figure 31) are likely to be uniquely identifiable by values of {*gender*, *date of birth*, *county*}. Wyoming had the largest percentage of its population identifiable (75%). Connecticut, Delaware, the District of Columbia and New Jersey had 0% identified. The average was 28% and the standard deviation was 22%.

The next to last row in Figure 31 labeled "USA" reports the results of applying the 3 steps of experiment J to all counties in the United States. As shown, 18.1% of the population of the United States is likely to be uniquely identified by values of {*gender*, *date of birth*, *county*}. The last row in Figure 31 labeled "%ID_pop" displays the percentage of people in each age subdivision who are likely to be uniquely identified by values of {*gender, date of birth*, *county*}. For example, it reports that 25.84% of the population of persons residing in the United States between the ages of 55 and 64 are likely to be uniquely identifiable based on {*gender*, *date of birth*, *county*}.

State	DOB %ID_pop	DOB Total	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
AL	31%	1,239,261	203,418	119,887	100,859	147,718	149,568	145,639	165,943	206,229
AK	43%	231,537	39,695	25,282	18,424	47,769	29,382	31,533	17,293	22,159
AZ	5%	168,352	23,995	13,659	8,351	15,873	23,248	23,557	26,589	33,080
AR	55%	1,286,703	204,611	126,862	85,675	165,982	171,952	153,203	149,635	228,783
CA	2%	482,182	74,362	42,716	33,536	62,826	50,565	61,321	70,890	85,966
CO	16%	530,181	94,650	50,001	38,332	85,809	86,915	60,219	46,794	67,461
CT	0%	-	-	-	-	-	-	-	-	-
DE	0%	-	-	-	-	-	-	-	-	-
DC	0%	-	-	-	-	-	-	-	-	-
FL	5%	680,438	109,084	74,526	59,719	96,106	93,589	80,169	64,489	102,756
GA	36%	2,335,158	385,475	236,121	182,875	311,416	297,509	294,860	257,992	368,910
HI	2%	24,302	-	4,985	3,356	-	20	5,039	4,127	6,775
ID	50%	504,176	84,045	54,338	27,716	64,270	70,098	61,874	63,762	78,073
IL	15%	1,733,651	294,307	164,151	119,585	237,212	225,134	210,334	189,098	293,830
IN	33%	1,805,518	310,118	183,259	129,393	268,623	228,630	204,738	205,590	275,167
IA	57%	1,574,848	267,585	153,138	102,462	208,798	216,811	168,181	161,950	295,923
KS	45%	1,117,968	187,792	105,602	71,548	150,530	142,864	128,522	120,241	210,869
KY	55%	2,015,672	339,649	199,166	162,837	287,814	264,521	239,055	215,956	306,674
LA	26%	1,103,759	166,000	99,616	76,791	129,159	151,255	132,897	154,279	193,762
ME	24%	289,549	45,914	25,233	25,821	33,214	34,481	37,839	34,151	52,896
MD	6%	288,043	36,084	19,602	20,508	36,667	32,605	29,983	51,224	61,370
MA	1%	30,080	2,997	1,179	914	2,739	3,651	8,515	7,345	2,740
MI	14%	1,270,356	187,954	101,271	84,202	147,645	144,700	167,106	186,504	250,974
MN	35%	1,545,738	264,233	169,559	98,392	209,122	217,857	169,995	152,615	263,965
MS	58%	1,503,027	258,287	160,447	109,981	202,539	185,987	179,021	161,836	244,929

Figure 30 Uniqueness of {*County*, *Gender*, *Date of birth*} respecting age distribution, part 1

The county having the largest population was Los Angeles County in California, with 8,863,164 people. The county having the smallest population was Yellowstone County in

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

Montana where only 52 people reside. The average population for a county is 79,182 and the standard deviation is 263,813. There are a total of 3,141 counties.

State	DOB %ID_pop	DOB Total	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
MO	35%	1,777,250	299,822	177,231	117,495	235,542	226,741	199,449	201,418	319,552
MT	58%	461,847	79,868	50,920	26,385	52,524	57,155	54,121	58,303	82,571
NE	52%	827,590	148,680	81,858	48,439	114,003	115,534	80,505	86,132	152,439
NV	17%	205,707	39,060	19,336	13,922	34,721	32,791	23,853	18,433	23,591
NH	14%	158,917	18,460	14,248	10,218	17,034	17,684	29,898	27,505	23,870
NJ	0%	13,203	-	0	-	-	-	7,089	6,114	-
NM	31%	466,942	55,935	41,285	37,298	48,782	59,912	68,042	65,578	90,110
NY	4%	714,072	86,136	44,198	48,241	39,884	79,329	130,425	139,808	146,051
NC	26%	1,718,318	242,446	149,044	126,104	202,459	209,646	232,090	249,455	307,074
ND	63%	401,471	65,977	37,393	19,272	49,281	47,612	47,773	53,295	80,868
OH	14%	1,536,542	244,518	135,966	102,380	185,611	200,338	199,829	190,210	277,690
OK	44%	1,395,889	214,447	135,344	106,030	180,771	170,655	171,464	164,825	252,353
OR	16%	468,933	58,089	37,189	25,490	51,118	50,034	77,348	76,465	93,200
PA	7%	868,774	143,074	81,634	65,841	120,867	115,691	110,104	109,679	121,884
RI	4%	36,592	7,442	4,146	-	-	7,157	5,220	4,929	7,698
SC	23%	792,897	115,127	71,978	53,250	100,618	97,592	104,465	111,669	138,198
SD	73%	506,465	96,431	52,085	32,146	70,960	65,236	51,201	50,256	88,150
TN	37%	1,832,875	296,158	180,822	134,829	239,766	227,196	226,279	223,498	304,327
TX	19%	3,185,236	555,868	314,582	220,496	408,489	396,535	357,970	372,889	558,407
UT	17%	296,513	58,729	33,397	21,901	43,107	40,697	30,917	26,743	41,022
VT	59%	329,450	48,194	35,514	24,136	42,551	42,821	44,422	36,277	55,535
VA	35%	2,186,920	327,643	195,729	180,037	286,163	280,550	300,469	262,255	354,074
WA	11%	523,874	66,444	57,010	36,219	58,899	62,605	75,825	83,264	83,608
WV	59%	1,059,753	168,623	100,661	72,214	144,775	151,174	140,705	128,935	152,666
WI	25%	1,211,247	190,779	110,977	71,189	162,807	159,047	141,883	157,036	217,529
WY	75%	338,752	57,064	36,055	20,545	52,035	52,251	38,218	35,539	47,045
USA	**18.1%**	45,076,528	7,265,269	4,329,202	3,175,354	5,854,598	5,787,325	5,543,164	5,448,813	7,672,803
%ID_pop			16.72%	18.27%	14.04%	13.48%	15.40%	21.79%	25.84%	24.65%

Figure 31 Uniqueness of {*County*, *Gender*, *Date of birth*} respecting age distribution, part 2

6. Discussion

Figure 32 contains a summary of the results reported in the previous section. A description of each reported percentage is provided in the following paragraphs. These percentages demonstrate how combinations of characteristics can combine to narrow the number of possible people under consideration as the subject of de-identified person-specific data.

County	18.1	0.04	0.00004	0.00000*
Place	58.4	3.6	0.04	0.01
ZIP	87.1	3.7	0.04	0.01
	DOB	**Mon/Year**	**BirthYear**	**2yr Age**

Figure 32 Percentage of US population identified with gender as geography and age vary

Experiment B reported that 87.1% (216 million of 248 million) of the population in the United States had characteristics that were likely made them unique based only on {*5-digit ZIP*, *gender*, *date of birth*}. Experiment C reported that 3.7% of the population in the United States had characteristics that were likely made them unique based only on {*5-digit ZIP*, *gender*, *Month and year of birth*}. Experiment D reported that 0.04% of the population in the United States had characteristics that were likely made them unique based only on {*5-digit ZIP*, *gender*, Year *of birth*}. Experiment E reported that 0.01% of the population in the United States had characteristics that were likely made them unique based only on {*5-digit ZIP*, *gender*, *2year age range*}.

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

Experiment F reported that 58.4% of the population in the United States had characteristics that were likely made them unique based only on {*Place*, *gender*, *date of birth*}. Experiment G reported that 3.6% of the population in the United States had characteristics that were likely made them unique based only on {*Place*, *gender*, *Month and year of birth*}. Experiment H reported that 0.04% of the population in the United States had characteristics that were likely made them unique based only on {*Place*, *gender*, *Year of birth*}. Experiment I reported that 0.01% of the population in the United States had characteristics that were likely made them unique based only on {*Place*, *gender*, *2year age range*}.

Experiment J reported that 18.1% of the population in the United States had characteristics that were likely made them unique based only on {*County*, *gender*, *date of birth*}. Experiment K reported that 0.04% of the population in the United States had characteristics that were likely made them unique based only on {*County*, *gender*, *Month and year of birth*}. Experiment L reported that 0.00004% of the population in the United States had characteristics that were likely made them unique based only on {*County*, *gender*, *Year of birth*}. Experiment M reported that 0.00000% of the population in the United States had characteristics that were likely made them unique based only on {*County*, *gender*, *2year age range*}, but despite it being a very small number, it is not 0.*

As the number of possible values a quasi-identifier can assume decreases, the percentage of the population in the United States who had characteristics that were likely unique based on those values decreases. This is evidenced by each row in Figure 32. Moving from left to right within each row of Figure 32, the numbers of possible combinations decrease and the corresponding percentages decrease. Aggregating the geographical specification to county resulted in far fewer possible combinations than available with place or ZIP codes. This is evidenced within each column in Figure 32. Notice however that the differences between the number of places and the number of ZIP codes are not as dramatic, and as a result, neither are the corresponding percentages.

6.1. Predicting the number of people that can be identified in a release

It was already shown that de-identified releases of person-specific data that contain no explicit identifiers such as name, address or phone number, is not necessarily anonymous [16]. The maximum number of patients who could be identified in a public or semi-public release of health data is the number of patients who were hospitalized and whose information is therefore included in the data. Many possible combinations of attributes can combine to form a quasi-identifier useful for linking the de-identified data to explicitly identified data. The number of hospitalizations reported in the IHCCCC's R_{rod} data (see Figure 2) in one year is estimated to be 1 million based on the average statistic that 1:12 people are hospitalized each year.

However, the actual number of patients that could be re-identified in publicly and semi-publicly released health data is not necessarily every patient and the actual number is likely to differ among releases due to varying quasi-identifiers available. The results from the experiments reported in this document can help predict a minimum level of identifiability based on a combination of three demographics.

* In Loving County, Texas, 6 of 107 people are likely to be uniquely identified by values of {*gender*, *2yr age range*, *county*}. All of these 6 people are between the ages of 12 and 18 years.

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

6.1.1. Illinois Research Health Data

As shown in Figure 2, R_{rod} includes the full date of birth, gender, and the patient's 5-digit residential ZIP. Figure 13 reports that 75.3% of the population of Illinois is likely to be uniquely identified by {*5-digit ZIP*, *gender*, *date of birth*}. That corresponds to 753,000 patients being identified per year in R_{rod}.

6.1.2. AHRQ's State Inpatient Database

As shown in Figure 3, SID includes the month and year of birth, gender, and the patient's 5-digit residential ZIP for some states. Figure 33 estimates that 112,595 patients per year are likely to be uniquely identified by {*ZIP*, *Gender*, *Month and year of birth*} in SID. The five states known to report the month and year of the birth date of each patient to SID were introduced in **Error! Reference source not found.**. The populations for each of these states according to the 1990 Census data [17] were reported in Figure 8 and Figure 9. It is estimated that 1:12 people are hospitalized each year. These values are summarized in Figure 33.

State	Population	Hospitalized	Unique	PopID
AZ	3,665,228	305,436	1.4%	4,276
IA	2,776,442	231,370	18.1%	41,878
NY	17,990,026	1,499,169	2.3%	34,481
OR	2,842,321	236,860	4.2%	9,948
WI	4,891,452	407,621	5.4%	22,012
		Total per year		***112,595***

Figure 33 Estimated Uniqueness of {*ZIP*, *Gender*, *Month and year of birth*} in SID

State	Population	Hospitalized	Unique	PopID
AZ	3,665,228	305,436	0.02%	61
CA	29,755,274	2,479,606	0.01%	248
CO	3,293,771	274,481	0.08%	220
FL	12,686,788	1,057,232	0.00%	42
IA	2,776,442	231,370	0.11%	255
MA	6,011,978	500,998	0.01%	50
MD	4,771,143	397,595	0.03%	119
NJ	7,730,188	644,182	0.01%	64
NY	17,990,026	1,499,169	0.03%	450
OR	2,842,321	236,860	0.07%	166
SC	3,486,703	290,559	0.01%	29
WA	4,866,692	405,558	0.03%	122
WI	4,891,452	407,621	0.02%	82
		Total per year		***1,907***

Figure 34 Estimated Uniqueness of {*ZIP*, *Gender*, *Year of birth*} in SID

As shown in Figure 3, SID includes the year of birth (by way of age[18]), gender, and the patient's 5-digit residential ZIP for some states. Figure 34 estimates that 1,907 patients per year are likely to be uniquely identified by {*ZIP*, *Gender*, *Year of birth*} in SID. The 13 states known to report the year of the birth date of each patient to SID were introduced in **Error! Reference source not found.**. The populations for each of these states according to the 1990 Census data [19] were reported in Figure 8 and Figure 9. It is estimated that 1:12 people are hospitalized each year. These values are summarized in Figure 34.

There are many ways to misunderstand these values. These values are not to be considered an estimate of the uniqueness of R_{rod} or SID. There may exist other quasi-identifiers that may consist of more and different attributes that can link to other available data and thereby render the released health data even more identifiable. Such quasi-identifiers may use the hospital identifying number or discharge status or payment information. The estimates reported in this document are just approximations based on the demographic quasi identifiers stated. Therefore, these estimates should be viewed as a minimal estimate of the identifiability of these data. Clearly, these data are not anonymous.

6.2. Unique and unusual information found in data

A significant problem with producing anonymous data concerns unique and unusual information appearing within the data themselves. Instances of uniquely occurring characteristics found within the original data can be used by a reporter, private investigator and others to discredit the anonymity of the released data even when these instances are not unique in the general population. Unusual cases are often unusual in other sources of data as well making them easier to identify.

Importantly, close examination of the particulars of a database provides the best basis for determining uniquely identifying information and quasi-identifiers. In this document, I have examined outside information without examining the values of the released data themselves. The analysis is based on the fact that a combination of characteristics that makes one unique in a geographic population, for example, results in uniqueness in all other data that includes that geographic specification. An examination of the data however can reveal other kinds of unusual information that can be found in other sources of data making more patients easier to identify.

In an interview, for example, a janitor may recall an Asian patient whose last name was Chan and who worked as a stockbroker because the patient gave the janitor some good investing tips. Any single uniquely occurring value or group of values can be used to identify an individual. Remember that the unique characteristic may not be known beforehand. It could be based on diagnosis, treatment, birth year, visit date, or some other little detail or combination of details available to the memory of a patient or a doctor, or knowledge about the database from some other source.

As another example, consider the medical records of a pediatric hospital in which only one patient is older than 45 years of age. Suppose a de-identified version of the hospital's records is to be released for public-use that includes age and city of residence but not birth date or zip code. Many may believe the resulting data would be anonymous because there are thousands of people of age 45 living in that city. However, the rare occurrence of a 45 year-old pediatric patient at that facility can become a focal point for anyone seeking to discredit the anonymity of the data. Nurses, clerks and other hospital personnel will often remember unusual cases and in interviews may provide additional details that help identify the patient.

6.3. Future Work

Below are proposed projects of varying degrees of difficulties and skill requirements that extend this work.

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

In this document, I have demonstrated how combinations of characteristics can combine to narrow the number of possible people under consideration. However, knowing that there exist a one or a few people that share particular characteristics and explicitly identifying those people are not exactly the same. These combinations of characteristics must be linked to explicitly identified information to reveal the identities of the individuals. Further demonstrate the identifiability of these data by providing population registers to which the data could be linked to re-identify the noted individuals.

In an earlier document [20], privacy risk measures were computed on the data sets R_{rod} and SID based on the assumption that the entire populations within those data were identifiable. While that may be correct, use the findings reported in this document, which are based only on basic demographic attributes and do not include other attributes within those data that could be used for re-identification, and re-compute the measures of risk for those collections. Make an argument as to why these re-computed risk measurements should be considered "minimal" risk values.

7. References

1 National Association of Health Data Organizations, *NAHDO Inventory of State-wide Hospital Discharge Data Activities* (Falls Church: National Association of Health Data Organizations, May 2000).

2 Cambridge Voters List Database. *City of Cambridge, Massachusetts*. Cambridge: February 1997.

3 Supra note 1 NAHDO.

4 State of Illinois Health Care Cost Containment Council, *Data release overview.* (Springfield: State of Illinois Health Care Cost Containment Council, March 1998).

5 Agency for Healthcare Research and Quality, *Healthcare Cost and Utilization Project: Central Distributor* (April, 2000) available at http://www.ahcpr.gov/data/hcup/hcup-pkt.htm.

6 1990 U.S. Census Data, Database C90STF3B. *U.S. Bureau of the Census*. Available at http://venus.census.gov and http://www.census.gov. Washington: 1993.

7 T. Dalenius. Finding a needle in a haystack – or identifying anonymous census record. *Journal of Official Statistics,* 2(3):329-336, 1986.

8 G. Smith. Modeling security-relevant data semantics. In *Proceedings of the 1990 IEEE Symposium on Research in Security and Privacy,* May 1990.

9 J. Ullman. *Principles of Database and Knowledge Base Systems*. Computer Science Press, Rockville, MD. 1988.

10 Supra note 6 U.S. Bureau of the Census.

11 “Census Counts 90,” *U.S. Bureau of the Census*. (Available on CDROM.) Washington: 1993.

12 "1996 National Five-Digit ZIP Code and Post Office Directory," *U.S. Postal Service*. Washington: 1996. Also available at http://www.usps.gov.

13 Brualdi, R.A., *Introductory Combinatorics*, North-Holland, New York, 1977.

14 Supra note 5 AHRQ.

15 L. Sweeney. *Inferences from unusual values in statistical data*. Carnegie Mellon Data Privacy Center Working Paper 3.

16 L. Sweeney. Weaving technology and policy together to maintain confidentiality. Journal of Law, Medicine and Ethics, 25(2-3):98-110, 1997.

17 Supra note 6 and note 11 U.S. Census.

18 Supra section 4.4.1 Age versus Year of Birth

19 Supra note 6 and note 11 U.S. Census.

20 L. Sweeney. *Towards all the data on all the people*. Formal publication forthcoming. Earlier version available as Carnegie Mellon Data Privacy Center Working Paper 2.

Exhibit 4

UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

January 19, 2018

MEMORANDUM FOR: Wilbur L. Ross, Jr.
Secretary of Commerce

Through: Karen Dunn Kelley
Performing the Non-Exclusive Functions and Duties of the Deputy Secretary

Ron S. Jarmin
Performing the Non-Exclusive Functions and Duties of the Director

Enrique Lamas
Performing the Non-Exclusive Functions and Duties of the Deputy Director

From: John M. Abowd
Chief Scientist and Associate Director for Research and Methodology

Subject: Technical Review of the Department of Justice Request to Add Citizenship Question to the 2020 Census

The Department of Justice has requested block-level citizen voting-age population estimates by OMB-approved race and ethnicity categories from the 2020 Census of Population and Housing. These estimates are currently provided in two related data products: the PL94-171 redistricting data, produced by April 1st of the year following a decennial census under the authority of 13 U.S.C. Section 141, and the Citizen Voting Age Population by Race and Ethnicity (CVAP) tables produced every February from the most recent five-year American Community Survey data. The PL94-171 data are released at the census block level. The CVAP data are released at the census block group level.

We consider three alternatives in response to the request: (A) no change in data collection, (B) adding a citizenship question to the 2020 Census, and (C) obtaining citizenship status from administrative records for the whole 2020 Census population.

We recommend either Alternative A or C. Alternative C best meets DoJ's stated uses, is comparatively far less costly than Alternative B, does not increase response burden, and does not harm the quality of the census count. Alternative A is not very costly and also does not harm the quality of the census count. Alternative B better addresses DoJ's stated uses than Alternative A. However, Alternative B is very costly, harms the quality of the census count, and would use substantially less accurate citizenship status data than are available from administrative sources.

Summary of Alternatives			
	Alternative A	***Alternative B***	***Alternative C***
Description	No change in data collection	Add citizenship question to the 2020 Census (i.e., the DoJ request), all 2020 Census microdata remain within the Census Bureau	Leave 2020 Census questionnaire as designed and add citizenship from administrative records, all 2020 Census microdata and any linked citizenship data remain within the Census Bureau
Impact on 2020 Census	None	Major potential quality and cost disruptions	None
Quality of Citizen Voting-Age Population Data	Status quo	Block-level data improved, but with serious quality issues remaining	Best option for block-level citizenship data, quality much improved
Other Advantages	Lowest cost alternative	Direct measure of self-reported citizenship for the whole population	Administrative citizenship records more accurate than self-reports, incremental cost is very likely to be less than $2M, USCIS data would permit record linkage for many more legal resident noncitizens
Shortcomings	Citizen voting-age population data remain the same or are improved by using small-area modeling methods	Citizenship status is misreported at a very high rate for noncitizens, citizenship status is missing at a high rate for citizens and noncitizens due to reduced self-response and increased item nonresponse, nonresponse followup costs increase by at least $27.5M, erroneous enumerations increase, whole-person census imputations increase	Citizenship variable integrated into 2020 Census microdata outside the production system, Memorandum of Understanding with United States Citizen and Immigration Services required to acquire most up-to-date naturalization data

Approved: _____ Date: _____

John M. Abowd, Chief Scientist
and Associate Director for Research and Methodology

DRAFT - deliberative

Detailed Analysis of Alternatives

The statistics in this memorandum have been released by the Census Bureau Disclosure Review Board with approval number CBDRB-2018-CDAR-014.

Alternative A: Make no changes

Under this alternative, we would not change the current 2020 Census questionnaire nor the planned publications from the 2020 Census and the American Community Survey (ACS). Under this alternative, the PL94-171 redistricting data and the citizen voting-age population (CVAP) data would be released on the current schedule and with the current specifications. The redistricting and CVAP data are used by the Department of Justice to enforce the Voting Rights Act. They are also used by state redistricting offices to draw congressional and legislative districts that conform to constitutional equal-population and Voting Rights Act nondiscrimination requirements. Because the block-group-level CVAP tables have associated margins of error, their use in combination with the much more precise block-level census counts in the redistricting data requires sophisticated modeling. For these purposes, most analysts and the DoJ use statistical modeling methods to produce the block-level eligible voter data that become one of the inputs to their processes.

If the DoJ requests the assistance of Census Bureau statistical experts in developing model-based statistical methods to better facilitate the DoJ's uses of these data in performing its Voting Rights Act duties, a small team of Census Bureau experts similar in size and capabilities to the teams used to provide the Voting Rights Act Section 203 language determinations would be deployed.

We estimate that this alternative would have no impact on the quality of the 2020 Census because there would be no change to any of the parameters underling the Secretary's revised life-cycle cost estimates. The estimated cost is about $350,000 because that is approximately the cost of resources that would be used to do the modeling for the DoJ.

Alternative B: Add the question on citizenship to the 2020 Census questionnaire

Under this alternative, we would add the ACS question on citizenship to the 2020 Census questionnaire and ISR instrument. We would then produce the block-level citizen voting-age population by race and ethnicity tables during the 2020 Census publication phase.

Since the question is already asked on the American Community Survey, we would accept the cognitive research and questionnaire testing from the ACS instead of independently retesting the citizenship question. This means that the cost of preparing the new question would be minimal. We did not prepare an estimate of the impact of adding the citizenship question on the cost of reprogramming the Internet Self-Response (ISR) instrument, revising the Census Questionnaire Assistance (CQA), or redesigning the printed questionnaire because those components will not be finalized until after the March 2018 submission of the final questions. Adding the citizenship question is similar in scope and cost to recasting the race and ethnicity questions again, should that become necessary, and would be done at the same time. After the 2020 Census ISR, CQA and printed questionnaire are in final form, adding the citizenship question would be much more expensive and would depend on exactly when the implementation decision was made during the production cycle.

For these reasons, we analyzed Alternative B in terms of its adverse impact on the rate of voluntary cooperation via self-response, the resulting increase in nonresponse followup (NRFU), and the consequent effects on the quality of the self-reported citizenship data. Three distinct analyses support the conclusion of an adverse impact on self-response and, as a result, on the accuracy and quality of the 2020 Census. We assess the costs of increased NRFU in light of the results of these analyses.

B.1. Quality of citizenship responses

We considered the quality of the citizenship responses on the ACS. In this analysis we estimated item nonresponse rates for the citizenship question on the ACS from 2013 through 2016. When item nonresponse occurs, the ACS edit and imputation modules are used to allocate an answer to replace the missing data item. This results in lower quality data because of the statistical errors in these allocation models. The analysis of the self-responses responses is done using ACS data from 2013-2016 because of operational changes in 2013, including the introduction of the ISR option and changes in the followup operations for mail-in questionnaires.

In the period from 2013 to 2016, item nonresponse rates for the citizenship question on the mail-in questionnaires for non-Hispanic whites (NHW) ranged from 6.0% to 6.3%, non-Hispanic blacks (NHB) ranged from 12.0% to 12.6%, and Hispanics ranged from 11.6 to 12.3%. In that same period, the ISR item nonresponse rates for citizenship were greater than those for mail-in questionnaires. In 2013, the item nonresponse rates for the citizenship variable on the ISR instrument were NHW: 6.2%, NHB: 12.3% and Hispanic: 13.0%. By 2016 the rates increased for NHB and especially Hispanics. They were NHW: 6.2%, NHB: 13.1%, and Hispanic: 15.5% (a 2.5 percentage point increase). Whether the response is by mail-in questionnaire or ISR instrument, item nonresponse rates for the citizenship question are much greater than the comparable rates for other demographic variables like sex, birthdate/age, and race/ethnicity (data not shown).

B.2. Self-response rate analyses

We directly compared the self-response rate in the 2000 Census for the short and long forms, separately for citizen and noncitizen households. In all cases, citizenship status of the individuals in the household was determined from administrative record sources, not from the response on the long form. A noncitizen household contains at least one noncitizen. Both citizen and noncitizen households have lower self-response rates on the long form compared to the short form; however, the decline in self-response for noncitizen households was 3.3 percentage points greater than the decline for citizen households. This analysis compared short and long form respondents, categories which were randomly assigned in the design of the 2000 Census.

We compared the self-response rates for the same household address on the 2010 Census and the 2010 American Community Survey, separately for citizen and noncitizen households. Again, all citizenship data were taken from administrative records, not the ACS, and noncitizen households contain at least one noncitizen resident. In this case, the randomization is over the selection of household addresses to receive the 2010 ACS. Because the ACS is an ongoing survey sampling fresh households each month, many of the residents of sampled households completed the 2010 ACS with the same reference address as they used for the 2010 Census. Once again, the self-response rates were lower in the ACS than in the 2010 Census for both citizen and noncitizen households. In this 2010 comparison, moreover, the decline in self-response was 5.1 percentage points greater for noncitizen households than for citizen households.

In both the 2000 and 2010 analyses, only the long-form or ACS questionnaire contained a citizenship question. Both the long form and the ACS questionnaires are more burdensome than the shortform. Survey methodologists consider burden to include both the direct time costs of responding and the indirect costs arising from nonresponse due to perceived sensitivity of the topic. There are, consequently, many explanations for the lower self-response rates among all household types on these longer questionnaires. However, the only difference between citizen and noncitizen households in our studies was the presence of at least one noncitizen in noncitizen households. It is therefore a reasonable inference that a question on citizenship would lead to some decline in overall self-response because it would make the 2020 Census modestly more burdensome in the direct sense, and potentially much more burdensome in the indirect sense that it would lead to a larger decline in self-response for noncitizen households.

B.3. Breakoff rate analysis

We examined the response breakoff paradata for the 2016 ACS. We looked at all breakoff screens on the ISR instrument, and specifically at the breakoffs that occurred on the screens with the citizenship and related questions like place of birth and year of entry to the U.S. Breakoff paradata isolate the point in answering the questionnaire where a respondent discontinues entering data—breaks off—rather than finishing. A breakoff is different from failure to self-respond. The respondent started the survey and was prepared to provide the data on the Internet Self-Response instrument, but changed his or her mind during the interview.

Hispanics and non-Hispanic non-whites (NHNW) have greater breakoff rates than non-Hispanic whites (NHW). In the 2016 ACS data, breakoffs were NHW: 9.5% of cases while NHNW: 14.1% and Hispanics: 17.6%. The paradata show the question on which the breakoff occurred. Only 0.04% of NHW broke off on the citizenship question, whereas NHNW broke off 0.27% and Hispanics broke off 0.36%. There are three related questions on immigrant status on the ACS: citizenship, place of birth, and year of entry to the United States. Considering all three questions Hispanics broke off on 1.6% of all ISR cases, NHNW: 1.2% and NHW: 0.5%. A breakoff on the ISR instrument can result in follow-up costs, imputation of missing data, or both. Because Hispanics and non-Hispanic non-whites breakoff much more often than non-Hispanic whites, especially on the citizenship-related questions, their survey response quality is differentially affected.

B.4. Cost analysis

Lower self-response rates would raise the cost of conducting the 2020 Census. We discuss those increased costs below. They also reduce the quality of the resulting data. Lower self-response rates degrade data quality because data obtained from NRFU have greater erroneous enumeration and whole-person imputation rates. An erroneous enumeration means a census person enumeration that should not have been counted for any of several reasons, such as, that the person (1) is a duplicate of a correct enumeration; (2) is inappropriate (e.g., the person died before Census Day); or (3) is enumerated in the wrong location for the relevant tabulation (https://www.census.gov/coverage_measurement/definitions/). A whole-person census imputation is a census microdata record for a person for which all characteristics are imputed.

Our analysis of the 2010 Census coverage errors (Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States, Memo G-01) contains the relevant data. That study found that when the 2010 Census obtained a valid self-response (219 million persons),

the correct enumeration rate was 97.3%, erroneous enumerations were 2.5%, and whole-person census imputations were 0.3%. All erroneous enumeration and whole-person imputation rates are much greater for responses collected in NRFU. The vast majority of NRFU responses to the 2010 Census (59 million persons) were collected in May. During that month, the rate of correct enumerations was only 90.2%, the rate of incorrect enumeration was 4.8%, and the rate of whole-person census imputations was 5.0%. June NRFU accounted for 15 million persons, of whom only 84.6% were correctly enumerated, with erroneous enumerations of 5.7%, and whole-person census imputations of 9.6%. (See Table 19 of 2010 Census Memorandum G-01. That table does not provide statistics for all NRFU cases in aggregate.)

One reason that the erroneous enumeration and whole-person imputation rates are so much greater during NRFU is that the data are much more likely to be collected from a proxy rather than a household member, and, when they do come from a household member, that person has less accurate information than self-responders. The correct enumeration rate for NRFU household member interviews is 93.4% (see Table 21 of 2010 Census Memorandum G-01), compared to 97.3% for non-NRFU households (see Table 19). The information for 21.0% of the persons whose data were collected during NRFU is based on proxy responses. For these 16 million persons, the correct enumeration rate is only 70.1%. Among proxy responses, erroneous enumerations are 6.7% and whole-person census imputations are 23.1% (see Table 21).

Using these data, we can develop a cautious estimate of the data quality consequences of adding the citizenship question. We assume that citizens are unaffected by the change and that an additional 5.1% of households with at least one noncitizen go into NRFU because they do not self-respond. We expect about 126 million occupied households in the 2020 Census. From the 2016 ACS, we estimate that 9.8% of all households contain at least one noncitizen. Combining these assumptions implies an additional 630,000 households in NRFU. If the NRFU data for those households have the same quality as the average NRFU data in the 2010 Census, then the result would be 139,000 fewer correct enumerations, of which 46,000 are additional erroneous enumerations and 93,000 are additional whole-person census imputations. This analysis assumes that, during the NRFU operations, a cooperative member of the household supplies data 79.0% of the time and 21.0% receive proxy responses. If all of these new NRFU cases go to proxy responses instead, the result would be 432,000 fewer correct enumerations, of which 67,000 are erroneous enumerations and 365,000 are whole-person census imputations.

For Alternative B, our estimate of the incremental cost proceeds as follows. Using the analysis in the paragraph above, the estimated NRFU workload will increase by approximately 630,000 households, or approximately 0.5 percentage points. We currently estimate that for each percentage point increase in NRFU, the cost of the 2020 Census increases by approximately $55 million. Accordingly, the addition of a question on citizenship could increase the cost of the 2020 Census by at least $27.5 million. It is worth stressing that this cost estimate is a lower bound. Our estimate of $55 million for each percentage point increase in NRFU is based on an average of three visits per household. We expect that many more of these noncitizen households would receive six NRFU visits.

We believe that $27.5 million is a conservative estimate because the other evidence cited in this report suggests that the differences between citizen and noncitizen response rates and data quality will be amplified during the 2020 Census compared to historical levels. Hence, the decrease in self-response for citizen households in 2020 could be much greater than the 5.1 percentage points we observed during the 2010 Census.

Alternative C: Use administrative data on citizenship instead of add the question to the 2020 Census

Under this alternative, we would add the capability to link an accurate, edited citizenship variable from administrative records to the final 2020 Census microdata files. We would then produce block-level tables of citizen voting age population by race and ethnicity during the publication phase of the 2020 Census using the enhanced 2020 Census microdata.

The Census Bureau has conducted tests of its ability to link administrative data to supplement the decennial census and the ACS since the 1990s. Administrative record studies were performed for the 1990, 2000 and 2010 Censuses. We discuss some of the implications of the 2010 study below. We have used administrative data extensively in the production of the economic censuses for decades. Administrative business data from multiple sources are a key component of the production Business Register, which provides the frames for the economic censuses, annual, quarterly, and monthly business surveys. Administrative business data are also directly tabulated in many of our products.

In support of the 2020 Census, we moved the administrative data linking facility for households and individuals from research to production. This means that the ability to integrate administrative data at the record level is already part of the 2020 Census production environment. In addition, we began regularly ingesting and loading administrative data from the Social Security Administration, Internal Revenue Service and other federal and state sources into the 2020 Census data systems. In assessing the expected quality and cost of Alternative C, we assume the availability of these record linkage systems and the associated administrative data during the 2020 Census production cycle.

C.1. Quality of administrate record versus self-report citizenship status

We performed a detailed study of the responses to the citizenship question compared to the administrative record citizenship variable for the 2000 Census, 2010 ACS and 2016 ACS. These analyses confirm that the vast majority of citizens, as determined by reliable federal administrative records that require proof of citizenship, correctly report their status when asked a survey question. These analyses also demonstrate that when the administrative record source indicates an individual is not a citizen, the self-report is "citizen" for no less than 23.8% of the cases, and often more than 30%.

For all of these analyses, we linked the Census Bureau's enhanced version of the SSA Numident data using the production individual record linkage system to append an administrative citizenship variable to the relevant census and ACS microdata. The Numident data contain information on every person who has ever been issued a Social Security Number or an Individual Taxpayer Identification Number. Since 1972, SSA has required proof of citizenship or legal resident alien status from applicants. We use this verified citizenship status as our administrative citizenship variable. Because noncitizens must interact with SSA if they become naturalized citizens, these data reflect current citizenship status albeit with a lag for some noncitizens.

For our analysis of the 2000 Census long-form data, we linked the 2002 version of the Census Numident data, which is the version closest to the April 1, 2000 Census date. For 92.3% of the 2000 Census long-form respondents, we successfully linked the administrative citizenship variable. The 7.7% of persons for whom the administrative data are missing is comparable to the item non-response for self-responders in the mail-in pre-ISR-option ACS. When the administrative data indicated that the 2000 Census respondent was a citizen, the self-response was citizen: 98.8%. For this same group, the long-form response was

noncitizen: 0.9% and missing: 0.3%. By contrast, when the administrative data indicated that the respondent was not a citizen, the self-report was citizen: 29.9%, noncitizen: 66.4%, and missing: 3.7%.

In the same analysis of 2000 Census data, we consider three categories of individuals: the reference person (the individual who completed the census form for the household), relatives of the reference person, and individuals unrelated to the reference person. When the administrative data show that the individual is a citizen, the reference person, relatives of the reference person, and nonrelatives of the reference person have self-reported citizenship status of 98.7%, 98.9% and 97.2%, respectively. On the other hand, when the administrative data report that the individual was a noncitizen, the long-form response was citizen for 32.9% of the reference persons; that is, reference persons who are not citizens according to the administrative data self-report that they are not citizens in only 63.3% of the long-form responses. When they are reporting for a relative who is not a citizen according to the administrative data, reference persons list that individual as a citizen in 28.6% of the long-form responses. When they are reporting for a nonrelative who is not a citizen according to the administrative data, reference persons list that individual as a citizen in 20.4% of the long-form responses.

We analyzed the 2010 and 2016 ACS citizenship responses using the same methodology. The 2010 ACS respondents were linked to the 2010 version of the Census Numident. The 2016 ACS respondents were linked to the 2016 Census Numident. In 2010, 8.5% of the respondents could not be linked, or had missing citizenship status on the administrative data. In 2016, 10.9% could not be linked or had missing administrative data. We reached the same conclusions using 2010 and 2016 ACS data with the following exceptions. When the administrative data report that the individual is a citizen, the self-response is citizen on 96.9% of the 2010 ACS questionnaires and 93.8% of the 2016 questionnaires. These lower self-reported citizenship rates are due to missing responses on the ACS, not misclassification. As we noted above, the item nonresponse rate for the citizenship question has been increasing. These item nonresponse data show that some citizens are not reporting their status on the ACS at all. In 2010 and 2016, individuals for whom the administrative data indicate noncitizen respond citizen in 32.7% and 34.7% of the ACS questionnaires, respectively. The rates of missing ACS citizenship response are also greater for individuals who are noncitizens in the administrative data (2010: 4.1%, 2016: 7.7%). The analysis of reference persons, relatives, and nonrelatives is qualitatively identical to the 2000 Census analysis.

In all three analyses, the results for racial and ethnic groups and for voting age individuals are similar to the results for the whole population with one important exception. If the administrative data indicate that the person is a citizen, the self-report is citizen at a very high rate with the remainder being predominately missing self-reports for all groups. If the administrative data indicate noncitizen, the self-report is citizen at a very high rate (never less than 23.8% for any racial, ethnic or voting age group in any year we studied). The exception is the missing data rate for Hispanics, who are missing administrative data about twice as often as non-Hispanic blacks and three times as often as non-Hispanic whites.

C.2. Analysis of coverage differences between administrative and survey citizenship data

Our analysis suggests that the ACS and 2000 long form survey data have more complete coverage of citizenship than administrative record data, but the relative advantage of the survey data is diminishing. Citizenship status is missing for 10.9 percent of persons in the 2016 administrative records, and it is missing for 6.3 percent of persons in the 2016 ACS. This 4.6 percentage point gap between administrative and survey missing data rates is smaller than the gap in 2000 (6.9 percentage points) and 2010 (5.6

percentage points). Incomplete (through November) pre-production ACS data indicate that citizenship item nonresponse has again increased in 2017.

There is an important caveat to the conclusion that survey-based citizenship data are more complete than administrative records, albeit less so now than in 2000. The methods used to adjust the ACS weights for survey nonresponse and to allocate citizenship status for item nonresponse assume that the predicted answers of the sampled non-respondents are statistically the same as those of respondents. Our analysis casts serious doubt on this assumption, suggesting that those who do not respond to either the entire ACS or the citizenship question on the ACS are not statistically similar to those who do; in particular, their responses to the citizenship question would not be well-predicted by the answers of those who did respond.

The consequences of missing citizenship data in the administrative records are asymmetric. In the Census Numident, citizenship data may be missing for older citizens who obtained SSNs before the 1972 requirement to verify citizenship, naturalized citizens who have not confirmed their naturalization to SSA, and noncitizens who do not have an SSN or ITIN. All three of these shortcomings are addressed by adding data from the United States Citizen and Immigration Services (USCIS). Those data would complement the Census Numident data for older citizens and update those data for naturalized citizens. A less obvious, but equally important benefit, is that they would permit record linkage for legal resident aliens by allowing the construction of a supplementary record linkage master list for such people, who are only in scope for the Numident if they apply for and receive an SSN or ITIN. Consequently, the administrative records citizenship data would most likely have both more accurate citizen status and fewer missing individuals than would be the case for any survey-based collection method. Finally, having two sources of administrative citizenship data permits a detailed verification of the accuracy of those sources as well.

C.3. Cost of administrative record data production

For Alternative C, we estimate that the incremental cost, except for new MOUs, is $450,000. This cost estimate includes the time to develop an MOU with USCIS, estimated ingestion and curation costs for USCIS data, incremental costs of other administrative data already in use in the 2020 Census but for which continued acquisition is now a requirement, and staff time to do the required statistical work for integration of the administrative-data citizenship status onto the 2020 Census microdata. This cost estimate is necessarily incomplete because we have not had adequate time to develop a draft MOU with USCIS, which is a requirement for getting a firm delivery cost estimate from the agency. Acquisition costs for other administrative data acquired or proposed for the 2020 Census varied from zero to $1.5M. Thus the realistic range of cost estimates, including the cost of USCIS data, is between $500,000 and $2.0M

Questions on the Jan 19 Draft Census Memo on the DoJ Citizenship Question Reinstatement Request

1. **With respect to Alternatives B and C, what is the difference, if any, between the time when the data collected under each alternative would be available to the public?**

   Since the collection of this data, whether from administrative records or from an enumerated question, occurs prior to the creation of the Microdata Detail File (MDF) from which all tabulations will be performed, there is no difference in the timing of when the data collected under either alternative B or C could be made available to the public. The exact date for completion of the MDF is still being determined as the 2020 Census schedule is matured. However, the 2020 Census is working towards publishing the first post-apportionment tabulation data products as early as the first week of February 2021.

2. **What is the "2020 Census publication phase" (page 1 of the Detailed Analysis for Alternative B) versus Alternative C? Would there be any difference?**

   The 2020 Census publication phase is a broad window stretching from the release of the apportionment counts by December 31, 2020 through the last data product or report published in FY 2023, the final year of decennial funding for the 2020 Census. However, as stated in the answer to question 1, these data could be made available to the public on the same schedule as any other post-apportionment tabulated data product regardless of whether alternative B or C is used in its collection.

3. **What is the non-response rate for: (A) each question on the 2000 and 2010 Decennial Census short form and (B) each question on the 2010 ACS and most recent ACS?**

   The table below shows the item non-response (INR) rate for each question on the 2000 and 2010 Decennial Census short form. This is the percentage of respondents who did not provide an answer to an item.

   Item Nonresponse Rates for 2000 and 2010 Short Form Person Questions

	Relationship	Sex	Age	Hispanic Origin	Race	Tenure
2010	1.5	1.5	3.5	3.9	3.3	4.5
2000	1.3	1.1	3.7	3.1	2.9	4.1

   Source: Rothhaas, Lestina and Hill (2012) Tables

   Notes and Soucre:
   Rothhaas, C., Lestina, F. and Hill, J. (2012) "2010 Decennial Census Item Nonresponse and Imputation Assessment Report" 2010 Census Program for Evaluations and Experiments, January 24, 2012.

From report:

The INR rate is essentially the proportion of missing responses before pre-editing or imputation procedures for a given item (i.e., the respondent did not provide an answer to the item). For INR, missing values are included in the rates, but inconsistent responses (i.e., incompatible with other responses) are considered non-missing responses.

Online link to 2010 report that has 2000 information as well.
https://www.census.gov/2010census/pdf/2010_Census_INR_Imputation_Assessment.pdf

See attached spreadsheet for the item allocation rates by questions for the ACS for 2010, 2013, and 2016.

4. **What was the total survey response rate (i.e., percentage of complete questionnaires) for the 2000 long form and the 2000 short form? Of the incomplete long forms, what percentage left the citizenship question blank? Of the completed long forms, what percentage (if known) contained incorrect responses to the citizenship question?**

We do not have measures of total survey response rates from the 2000 long form and 2000 short form available at this time. The mail response rate in 2000 was 66.4 percent for short forms and 53.9 percent for long forms. No analysis that we were aware of was conducted on the incomplete long forms that left the citizenship question blank. The Census 2000 Content Reinterview Survey showed low inconsistency of the responses to the citizenship question. Only 1.8 percent of the respondents changed answers in the reinterview.

Source for 2000 mail response rates:
https://www.census.gov/pred/www/rpts/A.7.a.pdf

Source for 2000 Content Reinterview Survey. Page 32 source.
https://www.census.gov/pred/www/rpts/B.5FR_RI.PDF

5. **For the 2000 long and short forms, what was the percentage unanswered (left blank) for each question (i.e., what percentage of the responses for each question (sex, race, ethnicity, income, citizenship, etc.) were left blank)?**

For the 2000 shortform, the table in question 3a provides the percentage unanswered for each question.

For the 2000 longform, Griffin, Love and Obenski (2003) summarized the Census 2000 longform responses. Allocation rates for individual items in Census 2000 were computed, but because of the magnitude of these data, summary allocation measures were derived.

These rates summarize completeness across all data items for occupied units (households) and are the ratio of all population and housing items that had values allocated to the total number of population and housing items required to have a response. These composite measures provide a summary picture of the completeness of all data. Fifty-four population items and 29 housing items are included in these summary measures. The analysis showed that 9.9 percent of the population question items and 12.5 percent of the housing unit question items required allocation. Allocation involves using statistical procedures, such as within-household or nearest neighbor matrices, to impute missing values.

https://ww2.amstat.org/sections/srms/Proceedings/y2003/Files/JSM2003-000596.pdf

6. **What was the incorrect response rate for the citizenship question that was asked on the Long Form during the 2000 Decennial Census? Does the response rate on the 2000 Long Form differ from the incorrect response rate on the citizenship question for the ACS?**

In the 2000 long form, 2.3 percent of persons have inconsistent answers, 89.4 percent have consistent answers, and 8.2 percent have missing citizenship data in the SSA Numident and/or the 2000 long form. Among persons with nonmissing citizenship data in the SSA Numident and/or the 2000 long form, 2.6 percent have inconsistent answers and 97.4 percent have consistent answers.

In the 2010 ACS, 3.1 percent of persons have inconsistent answers, 86.0 percent have consistent answers, and 10.8 percent have missing citizenship data in the SSA Numident and/or the 2010 ACS. Among persons with nonmissing citizenship data in the SSA Numident and/or the 2010 ACS, 3.6 percent have inconsistent answers and 96.4 percent have consistent answers.

In the 2016 ACS, 2.9 percent of persons have inconsistent answers, 81.2 percent have consistent answers, and 15.9 percent have missing citizenship data in the SSA Numident and/or the 2016 ACS. Among persons with nonmissing citizenship data in the SSA Numident and/or the 2016 ACS, 3.5 percent have inconsistent answers and 96.5 percent have consistent answers.

These ACS and 2000 Census long form rates are based on weighted data.

This shows that inconsistent response rates are higher in the 2010 and 2016 ACS than in the 2000 long form.

7. **What is the incorrect response rate on other Decennial or ACS questions for which Census has administrative records available (for example, age, sex or income)?**

Table 7a shows the agreement rates between the 2010 Census response and the SSA Numident for persons who could be linked and had nonmissing values, and Table 7b shows

the agreement rates between the 2010 ACS and the SSA Numident. Gender has low disagreement (0.4-0.5 percent), and white alone (0.9 percent), black alone (1.7-2 percent), and age (2.1 percent) also have low disagreement rates. Disagreement rates are greater for other races (e.g., 46.4-48.6 percent for American Indian or Alaska Native alone). Hispanic origin is not well measured in the Numident, because it contains a single race response, one of which is Hispanic.

Table 7a. Demographic Variable Agreement Rates Between the 2010 Census and the SSA Numident

2010 Census Response	Percent Agreement with SSA Numident
Hispanic	54.2
Not Hispanic	99.7
White Alone	99.1
Black Alone	98.3
American Indian or Alaska Native Alone	51.4
Asian Alone	84.3
Native Hawaiian or Other Pacific Islander Alone	74.4
Some Other Race Alone	17.7
Age	97.9
Gender	99.4

Source: Rastogi, Sonya, and Amy O'Hara, 2012, "2010 Census Match Study," 2010 Census Planning Memoranda Series No. 247.

Table 7b. Demographic Variable Agreement Rates Between the 2010 Census and the SSA Numident

2010 ACS Response	Percent Agreement with SSA Numident
White Alone	99.1
Black Alone	98.0
American Indian or Alaska Native Alone	53.6
Asian Alone	82.9
Native Hawaiian or Other Pacific Islander Alone	72.9
Some Other Race Alone	17.2
Age 0-2 Date of Birth	95.2
Age 3-17 Date of Birth	95.6
Age 18-24 Date of Birth	95.2
Age 25-44 Date of Birth	95.8
Age 45-64 Date of Birth	95.9
Age 65-74 Date of Birth	96.5
Age 75 and older Date of Birth	92.7
Male	99.5
Female	99.5

Source: Bhaskar, Renuka, Adela Luque, Sonya Rastogi, and James Noon, 2014, "Coverage and Agreement of Administrative Records and 2010 American Community Survey Demographic Data," CARRA Working Paper #2014-14.

Abowd and Stinson (2013) find correlations of 0.75-0.89 between Survey of Income and Program Participation (SIPP) and SSA Detailed Earnings Record annual earnings between 1990-1999.[1]

8. How does the Census presently handle responses on the (A) Decennial Census and (B) the ACS when administrative records available to the Census confirm that the response on the Decennial Census or ACS is incorrect? Is the present Census approach to incorrect responses based on practice/policy or law (statute or regulation)?

We have always based the short form Decennial Census and the ACS on self-response, and while we have procedures in place to address duplicate or fraudulent responses, we do not check the accuracy of the answers provided to the specific questions on the Census questionnaire. This is a long established practice at the Census Bureau that has been thoroughly tested and in place since 1970, when the Census Bureau moved to a mail-out/respond approach to the Decennial Census. Title 13 of the U.S. Code allows the Census Bureau to use alternative data sources, like administrative records, for a variety of purposes, and we are using data in new ways in the 2020 Census. While this includes the use of administrative records data to fill in areas where a respondent does not provide an answer, we have not explored the possibility of checking or changing responses that a responding household has provided in response to the questionnaire.

9. Please explain the differences between the self-response rate analysis and the breakoff rate analysis. The range of breakoff rates between groups was far smaller than the range of self-response rates between groups.

Self-response means that a household responded to the survey by mailing back a questionnaire or by internet, and a sufficient number of core questions were answered so that an additional field interview was not required.

A breakoff occurs when an internet respondent stops answering questions prior to the end of the questionnaire. In most cases the respondent answers the core questions before breaking off, and additional fieldwork is not required. The breakoff rates are calculated separately by which question screen was the last one reached before the respondent stopped answering altogether.

The share of Hispanic respondents who broke off at some point before the end of the questionnaire (17.6 percent) is much higher than for non-Hispanic whites (9.5 percent).

[1] Abowd, John M., and Martha H. Stinson, 2013, "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data," Review of Economics and Statistics, Vol. 95(55), pp. 1451-1467.

Spreading the overall breakoff rates over 134 screens in the questionnaire works out to quite small rates per screen. It works out to an average breakoff rate of 0.131 percent per screen for Hispanics and 0.066 percent for non-Hispanic whites.

10. The NRFU numbers are comparatively small – approximately one additional household for NRFU per Census enumerator. Is this really a significant source of concern?

Yes, this is a significant concern. First, it gives rise to incremental NRFU cost of at least $27.5 million. This is a lower bound becaues it assumes the households that do not self-respond because we added a question on citizenship have the same follow-up costs as an average U.S. household. They won't because these households overwhelmingly contain at least one noncitzen, and that is one of our acknowledged hard-to-count subpopulations.

11. Given that the breakoff rate difference was approximately 1 percent, why did Census choose to use the 5.1 percent number for assessing the cost of Alternative B?

If a household breaks off an internet response at the citizenship, place of birth, or year of entry screens, this means it would have already responded to the core questions. This would not trigger follow-up fieldwork and thus would not involve additional fieldwork costs. In contrast, if a household does not mail back a questionnaire or give an internet response, fieldwork will be necessary and additional costs will be incurred. Thus, the 5.1 percent number for differential self-response is more appropriate for estimating the additional fieldwork cost of adding a citizenship question.

12. Alternative C states that Census would use administrative data from the Social Security Administration, Internal Revenue Service, and "other federal and state sources." What are the other sources?

In addition to continuing the acquisition of the Social Security Administration and Internal Revenue Service data, the Census Bureau is in discussion with the U.S. Citizen and Immigration Services (USCIS) staff to acquire additional citizenship data.

13. Is Census confident that administrative data will be able to be used to determine citizenship for all persons (e.g., not all citizens have social security numbers)?

We are confident that Alternative C is viable and that we have already ingested enough high-quality citizenship administrative data from SSA and IRS. The USCIS data are not required. They would, however, make the citizenship voting age tabulations better, but the administrative data we've got are very good and better than the data from the 2000 Census and current ACS. The type of activities required for Alternative C already occur daily and routinely at the Census Bureau. We have been doing this for business data products,

including the Economic Censuses, for decades. We designed the 2020 Census to use this technology too.

14. For Alternative C, the memo says, "we assume the availability of these record linkage systems and associated administrative data" – does Census already have in place access to this data or would this need to be negotiated? If negotiated, for which data sets specifically?

The Census Bureau has longstanding contractual relationships with the Social Security Administration and the Internal Revenue Service that authorize the use of data for this project. For new data acquired for this project (i.e., USCIS) we would estimate a six-month development period to put a data acquisition agreement in place. That agreement would also include terms specifying the authorized use of data for this project.

15. Are there any privacy issues / sensitive information prohibitions that might prevent other agencies from providing such data?

There are no new privacy or sensitivity issues associated with other agencies providing citizenship data. We have received such information in the past from USCIS. We are currently authorized to receive and use the data from SSA and IRS that are discussed in Alternative C.

16. How long would Census expect any negotiation for access to data take? How likely is it that negotiations would be successful? Are MOA's needed/required?

Current data available to the Census Bureau provide the quality and authority to use that are required to support this project. Additional information potentially available from USCIS would serve to supplement/validate those existing data. We are in early discussions with USCIS to develop a data acquisition agreement and at this time have no indications that this acquisition would not be successful.

17. What limitations would exist in working with other agencies like IRS, Homeland Security, etc. to share data?

The context for sharing of data for this project is for a one-way sharing of data from these agencies to the Census Bureau. Secure file transfer protocols are in-place to ingest these data into our Title 13 protected systems. For those data already in-place at the Census Bureau to support this project, provisions for sharing included in the interagency agreement restrict the Census Bureau from sharing person-level microdata outside the Census Bureau's Title 13 protections. Aggregates that have been processed through the Bureau's disclosure avoidance procedures can be released for public use.

18. If Alternative C is selected, what is Census's backup plan if the administrative data cannot be completely collected and utilized as proposed?

The backup plan is to use all of the administrative data that we currently have, which is the same set that the analyses of Alternative C used. We have verified that this use is consistent with the existing MOUs. We would then use estimation and modeling techniques similar to those used for the Small Area Income and Poverty Estimates (SAIPE) to impute missing citizenship status for those persons for whom we do not have administrative records. These models would also include estimates of naturalizations that occurred since the administrative data were ingested.

19. Does Census have any reason to believe that access to existing data sets would be curtailed if Alternative C is pursued?

No we do not believe that any access to existing data sets would be curtailed if we pursue Alternative C.

20. Has the proposed Alternative C approach ever been tried before on other data collection projects, or is this an experimental approach? If this has been done before, what was the result and what were lessons learned?

The approach in Alternative C has been routinely used in processing the economic censuses for several decades. The Bureau's Business Register was specifically redesigned for the 2002 Economic Census in order to enhance the ingestion and use of administrative records from the IRS and other sources. The data in these administrative records are used to substitute for direct responses in the economic censuses for the unsampled entities. They are also used as part of the review, edit, and imputation systems for economic censuses and surveys. On the household side, the approach in Alternative C was used extensively to build the residential characteristics for OnTheMap and OnTheMap for Emergency Management.

21. Is using sample data and administrative records sufficient for DOJ's request?

The 2020 Census data combined with Alternative C are sufficient to meet DoJ's request. We do not anticipate using any ACS data under Alternative C.

22. Under Alternative C, If Census is able to secure interagency agreements to provide needed data sets, do we know how long it would take to receive the data transmission from other agencies and the length of time to integrate all that data, or is that unknown?

With the exception of the USCIS data, the data used for this project are already integrated into the 2020 Census production schema. In mid-to late 2018, we plan to acquire the USCIS data and with those data and our existing data begin to develop models and business rules to select citizenship status from the composite of sources and attach that characteristic to

each U.S. person. We expect the development and refinement of this process to continue into 2019 and to be completed by third quarter calendar year 2019.

23. Cross referencing Census decennial responses with numerous governmental data sets stored in various databases with differing formats and storage qualities sounds like it could be complicated. Does Census have an algorithm in place to efficiently combine and cross reference such large quantities of data coming from many different sources? What cost is associated with Alternative C, and what technology/plan does Census have in place to execute?

Yes, the 2018 Census End-to-End test will be implementing processing steps to be able to match Census responses to administrative record information from numerous governmental data sets. The Census Bureau has in place the Person Identification Validation System to assign Protected Identification Keys to 2020 Census responses. The required technology for linking in the administrative records is therefore part of the 2020 Census technology. This incremental cost factored into the estimate for Alternative C is for integrating the citizenship variable specifically, since that variable is not currently part of the 2020 Census design. No changes are required to the production Person Identification Validation system to integrate the administrative citizenship data.

24. For section C-1 of the memo, when did Census do the analyses of the incorrect response rates for non-citizen answers to the long form and ACS citizenship question? Were any of the analyses published?

The comparisons of ACS, 2000 Decennial Census longform and SSA Numident citizenship were conducted in January 2018. This analysis has not been published.

25. Has Census corrected the incorrect responses it found when examining non-citizen responses? If not, why not?

In the American Community Survey (ACS), and the short form Decennial Census, we do not change self-reported answers. The Decennial Census and the ACS are based on self-response and we accept the responses provided by households as they are given. While we have procedures in place to address duplicate or fraudulent responses, we do not check the accuracy of the answers provided to the specific questions on the Census questionnaires. This is a long established process at the Census Bureau that has been thoroughly tested and in place since 1970, when the Census Bureau moved to a mail-out/respond approach to the Decennial Census.

26. Has the Department of Justice ever been made aware of inaccurate reporting of ACS data on citizenship, so that they may take this into consideration when using the data?

Not exactly. The Census Bureau is in close, regular contact with the Department of Justice (DOJ) regarding their data requirements. Our counterparts at DOJ have a solid understanding of survey methodology and the quality of survey data, and they are aware of the public documentation on sampling and accuracy surrounding the ACS. However, the specific rate of accuracy regarding responses to the ACS question on citizenship has never been discussed.

27. Why has the number of persons who cannot be linked increased from 2010 to 2016?

The linkage between the ACS and administrative data from the SSA Numident and IRS ITIN tax filings depends on two factors: (a) the quality of the personally identifiable information (PII) on the ACS response and (b) whether the ACS respondent is in the SSN/ITIN universe.

With respect to the quality of the PII on the ACS, there may be insufficient information on the ACS due to item nonresponse or proxy response for the person to allow a successful match using the production record linkage system. There may also be more than one record in the Numident or ITIN IRS tax filings that matches the person's PII. Finally, there may be a discrepancy between the PII provided to the ACS and the PII in the administrative records.

Alternatively, the person may not be in the Numident or ITIN IRS tax filing databases because they are out of the universe for those administrative systems. This happens when the person is a citizen without an SSN, or when the person is a noncitizen who has not obtained an SSN or ITIN.

Very few of the unlinked cases are due to insufficient PII in the ACS or multiple matches with administrative records. The vast majority of unlinked ACS persons have sufficient PII, but fail to match any administrative records sufficiently closely. This means that most of the nonmatches are because the ACS respondent is not in the administrative record universe.

The incidence of ACS persons with sufficient PII but no match with administrative records increased between 2010 and 2016. One contributing factor is that the number of persons linked to ITIN IRS tax filings in 2016 was only 39 percent as large as in 2010, suggesting that either fewer of the noncitizens in the 2016 ACS had ITINs, or more of them provided PII in the ACS that was inconsistent with their PII in IRS records.

28. Independent of this memo, what action does Census plan to take in response to the analyses showing that non-citizens have been incorrectly responding to the citizenship question?

The Census Bureau does not have plans to make any changes to procedures in the ACS. However, we will continue to conduct thorough evaluations and review of census and survey data. The ACS is focusing our research on the potential use of administrative records

in the survey. For instance, we are exploring whether we can use IRS data on income to reduce the burden of asking questions on income on the ACS. We are concentrating initially on questions that are high burden, e.g., questions that are difficult to answer or questions that are seen as intrusive.

29. Did Census make recommendations the last time a question was added?

Since the short form Decennial Census was established in 2010, the only requests for new questions we have received have been for the ACS. And, in fact, requests for questions prior to 2010 were usually related to the Decennial Census Long Form. We always work collaboratively with Federal agencies that request a new question or a change to a question. The first step is to review the data needs and the legal justification for the new question or requested changes. If, through this process, we determine that the request is justified, we work with the other agencies to test the question (cognitive testing and field testing). We also work collaboratively on the analysis of the results from the test which inform the final recommendation about whether or not to make changes or add the question.

30. Does not answering truthfully have a separate data standard than not participating at all?

We're not sure what you're asking here. Please clarify the question.

31. What was the process that was used in the past to get questions added to the decennial Census or do we have something similar where a precedent was established?

Because no new questions have been added to the Decennial Census (for nearly 20 years), the Census Bureau did not feed bound by past precedent when considering the Department of Justices' request. Rather, the Census Bureau is working with all relevant stakeholders to ensure that legal and regulatory requirements are filled and that questions will produce quality, useful information for the nation. As you are aware, that process is ongoing at your direction.

32. Has another agency ever requested that a question be asked of the entire population in order to get block or individual level data?

Not to our knowledge. However, it is worth pointing out that prior to 1980 the short form of the Decennial Census included more than just the 10 questions that have been on the short form since 1990.

33. Would Census linking of its internal data sets, with other data sets from places like IRS and Homeland Security, have an impact on participation as well (i.e., privacy concerns)?

The potential that concerns about the use of administrative records could have an impact on participation has always been a concern of ours, and it's a risk that we're managing on our risk register. We've worked closely with the privacy community throughout the decade, and we established a working group on our National Advisory Committee to explore this issue. We've also regularly briefed the Congress about our plans. At this stage in the decade there does not appear to be extensive concerns among the general public about our approach to using administrative records in the Nonresponse Operation or otherwise. We will continue to monitor this issue.

34. Would Alternative C require any legislation? If so, what is the estimated time frame for approval of such legislation?

No.

35. Census publications and old decennial surveys available on the Census website show that citizenship questions were frequently asked of the entire population in the past. Citizenship is also a question on the ACS. What was the justification provided for citizenship questions on the (A) short form, (B) long form, and (C) ACS?

In 1940, the Census Bureau introduced the use of a short form to collect basic characteristics from all respondents, and a long form to collect more detailed questions from only a sample of respondents. Prior to 1940, census questions were asked of everyone, though in some cases only for those with certain characteristics. For example, in 1870, a citizenship question was asked, but only for respondents who were male and over the age of 21.

Beginning in 2005, all the long-form questions – including a question on citizenship -- were moved to the ACS. 2010 was the first time we conducted a short-form only census. The citizenship question is included in the ACS to fulfill the data requirements of the Department of Justice, as well as many other agencies including the Equal Employment Opportunities Commission, the Department of Health and Human Services, and the Social Security Administration.

ACS Item Allocation Rates for United States: 2016, 2013, 2010

Title	2016	2013	2010
Overall housing allocation rate occupied and vacant housing units	4.9	5.6	5.2
Overall person allocation rate total population	9.5	8.4	5.8
Vacancy status vacant housing units	3.9	3.5	2.9
Tenure occupied housing units	1.2	1.3	1.2
Units in structure occupied and vacant housing units	1.5	1.5	1.5
Year moved in occupied housing units	3	3	3.4
Month moved in occupied housing units into which households move in the last two years	0.7	0.7	0.7
Year built occupied and vacant housing units	18.2	17.1	16.2
Lot size occupied and vacant single family and mobile homes	3.9	3.9	4.2
Agricultural sales occupied and vacant single family and mobile homes with lot size greater than or equal to 1 acre	4	4.2	4.4
Business on property occupied and vacant single family and mobile homes	**	2.4	3
Number of rooms occupied and vacant housing units	5	5.5	5.2
Number of bedrooms occupied and vacant housing units	5.5	4.6	4.3
Running water occupied and vacant housing units	2.4	2.1	2
Flush toilet occupied and vacant housing units	**	2.2	2
Bathtub or shower occupied and vacant housing units	2.6	2.2	2
Sink with a faucet occupied and vacant housing units	2.6	2.2	2
Stove or range occupied and vacant housing units	3.1	2.8	2.5
Refrigerator occupied and vacant housing units	3.2	2.9	2.7
Telephone occupied housing units	1.5	1.2	1.1
Number of vehicles occupied housing units	1.2	1.4	1.3
Heating fuel occupied housing units	3.4	3.4	3.3
Monthly electricity cost occupied housing units	8.1	8.2	7.3
Monthly gas cost occupied housing units	9.6	9.9	9.8

Yearly water and sewer cost occupied housing units	8.5	8.8	8.1
Yearly other fuel cost occupied housing units	7.3	8.3	10.6
Yearly food stamp recipiency - household occupied housing units	1.7	1.7	1.3
Yearly real estate taxes owner-occupied housing units	16.7	18.5	16.3
Yearly property insurance owner-occupied housing units	23.9	25.6	23.2
Mortgage status owner-occupied housing units	2.2	2.5	2.1
Monthly mortgage payment owner-occupied housing units with a mortgage	10.5	12.4	10.7
Mortgage payment incl. real estate taxes owner-occupied housing units with a mortgage	6.2	6.9	(X)
Mortgage payment incl. insurance owner-occupied housing units with a mortgage	6.8	7.4	(X)
Second mortgage owner-occupied housing units	3.2	3.7	3.4
Home equity loan owner-occupied housing units	3.7	4.3	4.2
Other monthly mortgage payment(s) owner-occupied housing units with second mortgage or home equity loan	23.3	21.7	17.9
Property value owner-occupied housing units and vacant housing units for sale	11.6	12.9	12.3
Yearly mobile home costs occupied mobile homes and other units	21.7	21.5	19.9
Monthly condominium fee owner-occupied housing units	0.8	0.8	0.7
Monthly rent occupied housing units rented for cash rent and vacant housing units for rent	10.5	9.8	9.3
Meals included in rent occupied housing units rented for cash rent and vacant housing units for rent	2.1	2.1	2
Desktop/laptop/notebook computer occupied housing units	1.3	3.2	**
Handheld computer/smart mobile phone occupied housing units	**	3.3	**
Tablet or other portable wireless computer occupied housing units	1.6	**	**
Smartphone occupied housing units	1.6	**	**
Other computer occupied housing units	1.7	3.7	**
Household has internet access occupied housing units	3.3	4.4	**
Dial-up internet service occupied housing units with internet access	3.8	5.7	**
DSL internet service occupied housing units with internet access	**	5.7	**

Cable modem internet service occupied housing units with internet access	**	5.7	**
Fiber-optic internet service occupied housing units with internet access	**	5.7	**
Cellular data plan (formerly mobile broadband) occupied housing units with internet access	7.6	26.7	**
Satellite internet service occupied housing units with internet access	3.8	5.7	**
High speed internet service occupied housing units with internet accesss	3.8	**	**
Some other internet service occupied housing units with internet access	3.8	5.7	**
Race total population	1.5	1.6	1.5
Hispanic origin total population	1.8	2.1	1.8
Sex total population	0.1	0.1	0.1
Age total population	1.7	1.6	1.3
Relationship total household population	1.2	1.1	1.2
Marital status total population 15 years and over	5.3	4.8	3
Married past 12 months total population 15 years and over, except those never married	6.9	6.6	4.7
Widowed past 12 months total population 15 years and over, except those never married	7.4	7	4.5
Divorced past 12 months total population 15 years and over, except those never married	7.4	7	4.5
Times married total population 15 years and over, except those never married	8.1	7.8	5.1
Year last married total population 15 years and over, except those never married	13.5	13.3	11.4
Place of birth total population	9.1	8.6	6.5
Citizenship total population	6	5.2	2.7
Year of naturalization total population naturalized citizens	22.5	22.5	16.6
Year of entry total population not born in US	14.8	13.2	10.3
Speaks another language at home total population 5 years and over	6.8	5.9	3.4
Language spoken total population 5 years and over who speak another language at home	8.3	7	5.7
English ability total population 5 years and over who speak another language at home	7.1	5.9	4
School enrollment total population 3 years and over	6.7	6	3.7
Grade level attending			

total population 3 years and over enrolled	10.2	8.9	6
Educational attainment total population 3 years and over	8.5	8	5.6
Field of degree total population 25 years and over with a bachelor's degree or higher	13.5	12.4	9.8
Mobility status total population 1 years and over	7.2	6.5	4
Migration state/foreign county total population 1 years and over movers	13.2	11.3	7.1
Migration county total population 1 years and over movers within US	14.6	12.5	8.3
Migration minor civil division total population 1 years and over movers within US	14.2	12.1	8.4
Migration place total population 1 years and over movers within US	15	12.9	8.8
Health insurance thru employer/union total population	10.7	9	6.2
Health insurance purchased directly total population	11.3	9.7	6.9
Health insurance through Medicare total population	9.5	8.1	5.2
Health insurance through Medicaid total population	12.2	10.5	7.9
Health insurance through TRICARE total population	12.5	10.8	8.1
Health insurance through VA total population	12.3	10.7	8.1
Health ins. thru Indian Health Service total population	12.8	11.1	8.5
Visual difficulty total population	7.1	6.1	3.4
Hearing difficulty total population	6.8	5.9	3.2
Physical difficulty total population 5 years and over	7.5	6.7	3.5
Difficulty remembering total population 5 years and over	7.5	6.7	3.5
Difficulty dressing total population 5 years and over	7.5	6.7	3.5
Difficulty going out total population 16 years and over	7.3	6.5	3.4
Grandchildren living in home noninstitutionalized population 30 years and over	1.1	1	0.9
Responsibility for grandchildren noninstitutionalized population 30 years and over who are grandparents with grandchildren in the home	17.7	15.7	12
Months responsible for grandchildren noninstitutionalized population 30 years and over who are grandparents with grandchildren in the home that have responsibility	17.2	16.1	14.9
Fertility status female total population 15-50	7.8	6.7	3.7

Veteran status total population 17 years and over	7.3	6.8	3.8
Periods of military service total population 17 years and over on active duty now or previously	9.7	9.3	6.3
Service-connected disability rating total population 17 years and over, except those who never served in the Armed Forces	6.8	6.6	3.9
Service-connected disability rating value total population 17 years and over with a service-connected disability	0.2	0.2	0.7
Employment status recode noninstitutionalized population 16 years and over	8.7	8.1	5.1
When last worked noninstitutionalized population 16 years and over	9.6	9.1	5.7
Weeks worked in the past 12 months noninstitutionalized population 16 years and over who worked in the past 12 months	10.6	9.7	6.9
Hours worked per week noninstitutionalized population 16 years and over who worked in the past 12 months	11.9	10.8	7.7
Place of work state/foreign county noninstitutionalized population 16 years and over at work last week	11.8	10.4	6.3
Place of work county noninstitutionalized population 16 years and over at work last week	12.5	11	7
Place of work minor civil division noninstitutionalized population 16 years and over at work last week	3.6	3.3	2.1
Place of work place noninstitutionalized population 16 years and over at work last week	13.1	11.6	7.6
Transportation to work noninstitutionalized population 16 years and over at work last week	9.6	8.8	5.7
Carpool size noninstitutionalized population 16 years and over at work last week who drive to work	10.9	9.9	6.8
Time of departure noninstitutionalized population 16 years and over at work last week who don't work at home	20.2	18.5	12.8
Commuting time noninstitutionalized population 16 years and over at work last week who don't work at home	14.5	13.3	9.7
Class of worker total population 16 years and over who worked in the last 5 years	11.7	10.7	7.2
Industry total population 16 years and over who worked in the last 5 years	12.7	11.4	7.8
Occupation total population 16 years and over who worked in the last 5 years	13.4	11.8	8.1
Wages/salary income total population 15 years and over	19.1	19	16
Self-employment income			

total population 15 years and over	10.5	9.3	5.9
Interest, dividends, etc. income			
total population 15 years and over	15.2	12.6	8.8
Social security or railroad retirement			
total population 15 years and over	14.5	12.3	8.9
Supplemental security income			
total population 15 years and over	12.7	10.3	6.7
Public assistance			
total population 15 years and over	13.2	10.5	6.8
Retirement income			
total population 15 years and over	13.6	11.1	7.5
Other income			
total population 15 years and over	13.2	10.8	7.4
Some or all income allocated			
total population 15 years and over	28.4	25.3	22.4

Source: ACS 1-year data. See following links for more information:

https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/item-allocation-rates/

https://www.census.gov/programs-surveys/acs/methodology/sample-size-and-data-quality/item-allocation-rates-definitions.html

Note:

** This item was not asked in this year.

Summary Analysis of the Key Differences Between Alternative C and Alternative D

This short note describes the Census Bureau's current assumptions about two alternatives to address the need for block level data on citizen voting age populations. The goal is to measure the citizenship status of all people enumerated in the 2020 Decennial Census. Both alternatives utilize administrative data on the citizenship status of individuals, however one option, Alternative D, proposes to also include the current American Community Survey (ACS) question on citizenship status on the 2020 Decennial Census short form.

In both alternatives described here, the methodology requires linking 2020 census response data and administrative records. However, as illustrated both alternatives would also need to assign/impute citizenship for a portion of the population. The Census Bureau will have to assign citizenship in cases of questionnaire non-response and item non-response. Additionally, it is important to note, that even when a self-response is available it is not always possible to link response data with administrative records data. Poor data quality (e.g., name and age) and nonresponse or incomplete 2020 Census responses mean that we will not have a direct measure of citizenship status for all residents enumerated in 2020. The Census Bureau will to need employ an imputation model for these cases.

One of the key differences between to the two alternatives described below is the number of cases requiring imputation. The other key difference is the impact of errors in the citizenship status reported on the 2020 Census.

In the most recent version of the 2020 Decennial Life Cycle Cost Estimate, the Census Bureau projects counting 330 million residents in 2020. Figure 1 summarizes how citizenship status will be measured under Alternative C that does not employ a citizenship question on the 2020 Census. Figure 2 summarizes how this will be done using both administrative records and a 2020 citizenship question under Alternative D.

Alternative C is a simplified process for assigning citizenship through direct linkage and modelling, without including the question on the 2020 Census. The Census Bureau will link the responses for the 330 million census records to administrative records that contain information on the citizenship status of individuals. The Census Bureau expects to successfully link and observe this status for approximately 295 million people. The Census Bureau would need to impute this status for approximately 35 million people under Alternative C whose 2020 responses cannot be linked to administrative data. Although the Census Bureau has fully developed and tested the imputation model, it has high confidence that an accurate model can be developed and deployed for this purpose. Further, we will most likely never possess a fully adequate truth deck to benchmark it to.

Measuring citizenship status is slightly more complex under Alternative D where all U.S. households will be given the opportunity to provide the citizenship status of each household member. Based on response data for the ACS citizenship and other response data research, we know that not all households that respond to the 2020 Census will answer this question, leaving the question blank or with otherwise invalid responses. Additionally, Alternative D, must also account for those households that do not respond at all or will have proxy responses. Due to these reasons, we estimate that we will get 2020 citizenship status responses for approximately 294.6 million people, a slightly higher estimate

than Alternative C. For the 35.4 million people without a 2020 citizenship response, the Census Bureau will employ the same methodology as in Alternative C, linking the 2020 Census responses to the administrative records. The Census Bureau estimates that it will be able to link these cases to administrative records where we observe citizenship status for approximately 21.5 million people. For the remaining 13.8 million will be imputed through a model as described above. Thus, there will be a need for imputing many cases across either alternative.

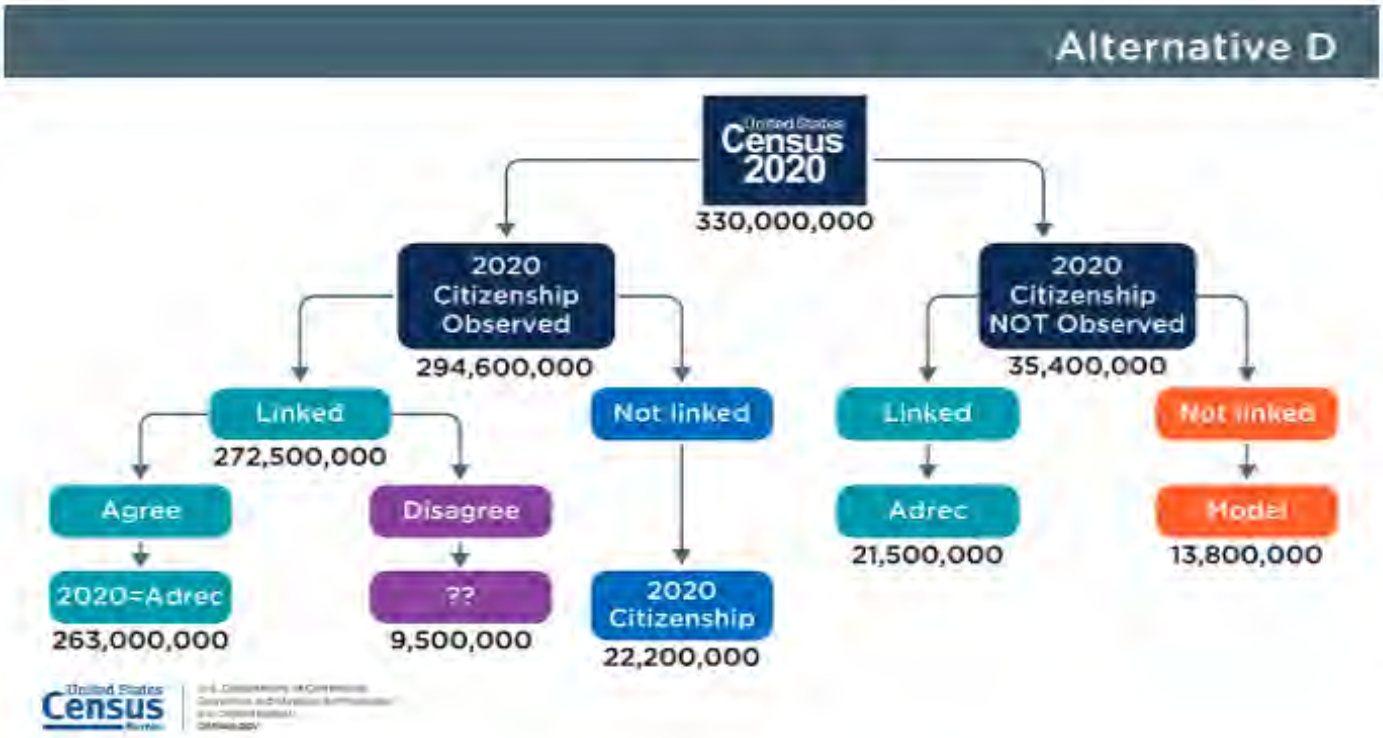
The Census Bureau will link the 294.6 million records from the 2020 Census with the administrative records. This will be done both for potential quality assurance purposes and to improve the quality of future modeling uses. Based on the current research from the ACS, the Census Bureau expects to successfully link approximately 272.5 million of these cases. Of these, 263 million will have citizenship statuses that agree across the 2020 response and administrative record. The Census Bureau estimates there will be 9.5 million cases where there is disagreement across the two sources. Historic Census Bureau practice is to use self-reported data in these situations. However, the Census Bureau now knows from linking ACS responses on citizenship to administrative data that nearly one third of noncitizens in the administrative data respond to the questionnaire indicating they are citizens, indicating that this practice should be revisited in the case of measuring citizenship. Finally, for those 22.2 million cases that do not link to administrative records (non-linkage occurs for the same data quality reasons discussed above), the Census Bureau will use the observed 2020 responses. Again, Census Bureau expect some quality issues with these responses. Namely, the Census Bureau estimates that just under 500 thousand noncitizens will respond as citizens.

The relative quality of Alternative C versus Alternative D will depend on the relative importance of the errors in administrative data, response data, and imputations. To be slightly more but not fully precise consider the following description of errors under both alternatives. First note that all possible measurement methods will have errors. Under Alternative C, there will be error in the administrative records, but we believe these to be relatively limited dues to the procedure following by SSA, USCIS and State. In both Alternative, the modeled cases will be subject to prediction error. Prediction error occur when the model returns the incorrect status of a case. As there are more models cases in Alternative C, prediction error will be a bigger issue there. Alternative D has an additional source or error, response error. This is where 2020 respondent give the incorrect status. Statisticians often hope these error are random and cancel out. However, we know from prior research that citizenship status responses are systematically biased for a subset of noncitizens. Response error is only an issue in alternative D. Unfortunately, the Census Bureau cannot quantify the relative magnitude of the errors across the alternatives at this time.

Figure 1

Alternative C
United States Census 2020
330,000,000
Linked
Not linked
Adrec
Model
295,000,000
35,000,000
United States Census Bureau

Figure 2


Alternative D
United States Census 2020
330,000,000
2020 Citizenship Observed
294,600,000
2020 Citizenship NOT Observed
35,400,000
Linked
272,500,000
Not linked
Linked
Not linked
Agree
Disagree
Adrec
Model
21,500,000
13,800,000
2020=Adrec
263,000,000
??
9,500,000
2020 Citizenship
22,200,000
United States Census Bureau

UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

March 1, 2018

MEMORANDUM FOR: Wilbur L. Ross, Jr.
Secretary of Commerce

Through: Karen Dunn Kelley
Performing the Non-Exclusive Functions and Duties of the Deputy Secretary

Ron S. Jarmin
Performing the Non-Exclusive Functions and Duties of the Director

Enrique Lamas
Performing the Non-Exclusive Functions and Duties of the Deputy Director

From: John M. Abowd
Chief Scientist and Associate Director for Research and Methodology

Subject: Preliminary analysis of Alternative D (Combined Alternatives B and C)

See attached.

Approved: _____ Date: _____
John M. Abowd, Chief Scientist
and Associate Director for Research and Methodology

DRAFT Pre-decisional (V1.0)

Preliminary Analysis of Alternative D

At the Secretary's request we performed a preliminary analysis of combining Alternative B (asking the citizenship question of every household on the 2020 Census) and Alternative C (do not ask the question, link reliable administrative data on citizenship status instead) in the January 19, 2018 draft memo to the Department of Commerce into a new Alternative D. Here we discuss Alternative D, the weaknesses in Alternative C on its own, whether and how survey data could address these weaknesses, implications of including a citizenship question for using administrative data, and methodological challenges.

Description of Alternative D: Administrative data from the Social Security Administration (SSA), Internal Revenue Service (IRS), U.S. Citizenship and Immigration Services (USCIS), and the State Department would be used to create a comprehensive statistical reference list of current U.S. citizens. Nevertheless, there will be some persons for whom no administrative data are available. To obtain citizenship information for this sub-population, a citizenship question would be added to the 2020 Census questionnaire. The combined administrative record and 2020 Census data would be used to produce baseline citizenship statistics by 2021. Any U.S. citizens appearing in administrative data after the version created for the 2020 Census would be added to the comprehensive statistical reference list. There would be no plan to include a citizenship question on future Decennial Censuses or American Community Surveys. The comprehensive statistical reference list, built from administrative records and augmented by the 2020 Census answers would be used instead. The comprehensive statistical reference list would be kept current, gradually replacing almost all respondent-provided data with verified citizenship status data.

What are the weaknesses in Alternative C?

In the 2017 Numident (the latest available), 6.6 million persons born outside the U.S. have blank citizenship among those born in 1920 or later with no year of death. The evidence suggests that citizenship is not missing at random. Of those with missing citizenship in the Numident, a much higher share appears to be U.S. citizens than compared to those for whom citizenship data are not missing. Nevertheless, some of the blanks may be noncitizens, and it would thus be useful to have other sources for them.

A second question about the Numident citizenship variable is how complete and timely its updates are for naturalizations. Naturalized citizens are instructed to immediately apply for a new SSN card. Those who wish to work have an incentive to do so quickly, since having an SSN card with U.S. citizenship will make it easier to pass the E-Verify process when applying for a job, and it will make them eligible for government programs. But we do not know what fraction of naturalized citizens actually notify the SSA, and how soon after being naturalized they do so.

A third potential weakness of Numident citizenship is that some people are not required to have a Social Security Number (SSN), whether they are a U.S. citizen or not. It would also be useful to have a data source on citizenship that did not depend on the SSN application and tracking process inside SSA. This is why we proposed the MOU with the USCIS for naturalizations, and why we have now begun pursuing an MOU with the State Department for data on all citizens with passports.

IRS Individual Taxpayer Identification Numbers (ITIN) partially fill the gap in Numident coverage of noncitizen U.S. residents. However, not all noncitizen residents without SSNs apply for ITINs. Only those making IRS tax filings apply for ITINs. Once again, it would be useful to have a data source that did not depend on the ITIN process. The USCIS and State Department MOUs would provide an alternative source in this context as well.

U.S. Citizenship and Immigration Services (USCIS) data on naturalizations, lawful permanent residents, and I-539 non-immigrant visa extensions can partially address the weaknesses of the Numident. The USCIS data provide up-to-date information since 2001 (and possibly back to 1988, but with incomplete records prior to 2001). This will fill gaps for naturalized citizens, lawful permanent residents, and persons with extended visa applications without SSNs, as well as naturalized citizens who did not inform SSA about their naturalization. The data do not cover naturalizations occurring before 1988, as well as not covering and some between 1988-2000. USCIS data do not always cover children under 18 at the time a parent became a naturalized U.S. citizen. Such children automatically become U.S. citizens under the Child Citizenship Act of 2000. The USCIS receives notification of some, but not all, of these child naturalizations. Others inform the U.S. government of their U.S. citizenship status by applying for U.S. passports, which are less expensive than the application to notify the USCIS. USCIS visa applications list people's children, but those data may not be in electronic form.

U.S. passport data, available from the State Department, can help plug the gaps for child naturalizations, blanks on the Numident, and out-of-date citizenship information on the Numident for persons naturalized prior to 2001. Since U.S. citizens are not required to have a passport, however, these data will also have gaps in coverage.

Remaining citizenship data gaps in Alternative C include the following categories:

1. U.S. citizens from birth with no SSN or U.S. passport. They will not be processed by the production record linkage system used for the 2020 Census because their personally identifiable information won't find a matching Protected Identification Key (PIK) in the Person Validation System (PVS).

2. U.S. citizens from birth born outside the U.S., who do not have a U.S. passport, and either applied for an SSN prior to 1974 and were 18 or older, or applied before the age of 18 prior to 1978. These people will be found in PVS, but none of the administrative sources discussed above will reliably generate a U.S. citizenship variable.

3. U.S. citizens who were naturalized prior to 2001 and did not inform SSA of their naturalization because they originally applied for an SSN after they were naturalized, and it was prior to when citizenship verification was required for those born outside the U.S. (1974). These people already had an SSN when they were naturalized and they didn't inform SSA about the naturalization, or they didn't apply for an SSN. The former group have inaccurate data on the Numident. The latter group will not be found in PVS.

4. U.S. citizens who were automatically naturalized if they were under the age of 18 when their parents became naturalized in 2000 or later, and did not inform USCIS or receive a U.S. passport. Note that such persons would not be able to get an SSN with U.S. citizenship on the card without either a U.S. passport or a certificate from USCIS. These people will also not be found in the PVS.

5. Lawful permanent residents (LPR) who received that status prior to 2001 and either do not have an SSN or applied for an SSN prior to when citizenship verification was required for those born outside the U.S. (1974). The former group will not be found in PVS. The latter group has inaccurate data in Numident.

6. Noncitizen, non-LPR, residents who do not have an SSN or ITIN and who did not apply for a visa extension. These persons will not be found in PVS.

7. Persons with citizenship information in administrative data, but the administrative and decennial census data cannot be linked due to missing or discrepant PII.

Can survey data address the gaps in Alternative C?

One might think that survey data could help fill the above gaps, either when their person record is not linked in the PVS, and thus they have no PIK, or when they have a PIK but the administrative data lack up-to-date citizenship information. Persons in Category 6, however, have a strong incentive to provide an incorrect answer, if they answer at all. A significant, but unknown, fraction of persons without PIKs are in Category 6. Distinguishing these people from the other categories of persons without PIKs is an inexact science because there is no feasible method of independently verifying their non-citizen status. Our comparison of ACS and Numident citizenship data suggests that a large fraction of LPRs provide incorrect survey responses. This suggests that survey-collected citizenship data may not be reliable for many of the people falling in the gaps in administrative data. This calls into question their ability to improve upon Alternative C.

With Alternative C, and no direct survey response, the Census Bureau's edit and imputation procedures would make an allocation based primarily on the high-quality administrative data. In the presence of a survey response, but without any linked administrative data for that person, the edit would only be triggered by blank citizenship. A survey response of "citizen" would be accepted as valid. There is no scientifically defensible method for rejecting a survey response in the absence of alternative data for that respondent.

How might inclusion of a citizenship question on the questionnaire affect the measurement of citizenship with administrative data? Absent an in-house administrative data census, measuring citizenship with administrative data requires that persons in the Decennial Census be linked to the administrative data at the person level. The PVS system engineered into the 2020 Census does this using a very reliable technology. However, inclusion of a citizenship question on the 2020 Census questionnaire is very likely to reduce the self-response rate, pushing more households into Nonresponse Followup (NRFU). Not only will this likely lead to more incorrect enumerations, but it is also expected to increase the number of persons who cannot be linked to the administrative data because the NRFU PII is lower quality than the self-response data. In the 2010 Decennial Census, the percentage of NRFU persons who could be linked to administrative data rate was 81.6 percent, compared to 96.7 percent for mail responses. Those refusing to self-respond due to the citizenship question are particularly likely to refuse to respond in NRFU as well, resulting in a proxy response. The NRFU linkage rates were far lower for proxy responses than self-responses (33.8 percent vs. 93.0 percent, respectively).

Although persons in Category 6 will not be linked regardless of response mode, it is common for households to include persons with a variety of citizenship statuses. If the whole household does not self-

respond to protect the members in Category 6, the record linkage problem will be further aggravated. Thus, not only are citizenship survey data of suspect quality for persons in the gaps for Alternative C, collecting these survey data would reduce the quality of the administrative records when used in Alternative D by lowering the record linkage rate for persons with administrative citizenship data.

What methodological challenges are involved when combining these sources?

Using the 2020 Census data only to fill in gaps for persons without administrative data on citizenship would raise questions about why 100 percent of respondents are being burdened by a citizenship question to obtain information for the two percent of respondents where it is missing.

Including a citizenship question in the 2020 Census does not solve the problem of incomplete person linkages when producing citizenship statistics after 2020. Both the 2020 decennial record and the record with the person's future location would need to be found in PVS to be used for future statistics.

In sum, Alternative D would result in poorer quality citizenship data than Alternative C. It would still have all the negative cost and quality implications of Alternative B outlined in the draft January 19, 2018 memo to the Department of Commerce.

DRAFT Pre-decisional (v. 0)