2020 CENSUS PROGRAM INTERNAL MEMORANDUM SERIES: 2019.23.i

Date: May 31, 2019

MEMORANDUM FOR: The Record

From: Deborah M. Stempowski **(Signed May 31, 2019)**
                       Chief, Decennial Census Management Division

Subject: 2020 Census Evaluation: Evaluating Privacy and Confidentiality Concerns Study Plan

Contact: Jennifer Reichert
                       Decennial Census Management Division
                       301-763-4298
                       jennifer.w.reichert@census.gov

This memorandum releases the final version of the 2020 Census Evaluation: Evaluating Privacy and Confidentiality Concerns Study Plan, which is part of the 2020 Census Program for Evaluations and Experiments (CPEX). For specific content related questions, you may also contact the authors:

Aleia Clark Fobia
Center for Behavioral Science Methods
301-763-4075
aleia.yvonne.clark.fobia@census.gov

Jennifer Hunter Childs
Center for Behavioral Science Methods
301-763-4927
jennifer.hunter.childs@census.gov

Matthew Virgile
Center for Behavioral Science Methods
301-763-4745
matthew.virgile@census.gov

Jessica Holzberg
Center for Behavioral Science Methods
301-763-2298
jessica.holzberg@census.gov

Kelly Mathews
Decennial Statistical Studies Division
301-763-5639
kelly.m.mathews@census.gov

2020 Census Evaluation

Evaluating Privacy and Confidentiality Concerns Study Plan

Aleia Clark Fobia, CBSM
Jessica Holzberg, CBSM
Jennifer Hunter Childs, CBSM
Kelly Mathews, DSSD
Matthew Virgile, CBSM

Page intentionally left blank.

Table of Contents

Page intentionally left blank.

I.   Introduction

Privacy and confidentiality has been at the forefront of concerns as the census moves online and increases reliance on administrative records. The Census Bureau has been conducting research on respondents' privacy and confidentiality concerns with online response and administrative records use as the focus of one of the teams from the Research and Testing phase leading up to the 2020 Census. Thus far, much of this work has been hypothetical, with respondents asked how they would feel if x strategy were to be employed in the census. The 2020 Census provides an opportunity to evaluate respondent privacy and confidentiality concerns and their relationship to response mode, item nonresponse, and mismatches between administrative records and self-reported data in a decennial census environment.

Privacy and confidentiality research addresses key elements in the 2020 Census Operational Plan and the guiding principles for the 2030 Census. As the 2020 Census is the first time that the majority of respondents will be encouraged to respond to the census on the internet, a key element of the Optimizing Self-Response innovation area is assuring respondents that their data is secure and treated as confidential (US Census Bureau, 2017; p. 19). Government and private sector data breaches are salient public events that have potentially weakened respondents' trust in the Census Bureau's ability to maintain privacy and keep data confidential. Respondents need to be sure that their personal information is protected, particularly when responding online. In fact, the public's perception of the Census Bureau's ability to safeguard response data has been identified as a high-level risk to the 2020 Census Program (Blumerman & Fontenot, 2017).

This evaluation is a telephone and in-person survey of decennial census respondents focused on their privacy and confidentiality concerns. The evaluation is an opportunity to measure how the web response option affects privacy and confidentiality concerns of respondents who have had the chance to use this option. Based on previous research, we expect that respondents will have particular privacy and confidentiality concerns associated with responding online (Holzberg & Fobia, 2016; Morales, Holzberg, & Eggleston, 2017). The public perception of how the Census Bureau handles privacy and confidentiality in 2020 will shape how the Census Bureau prepares for and executes a 2030 Census, especially one that might be all-electronic.

Expanded use of administrative records in 2020 and the principle of a primarily records-based census in 2030 are also areas where research about respondent privacy and confidentiality attitudes is crucial. The 2020 Census plans to use administrative records and third-party data to target advertising, validate respondent submissions, and reduce nonresponse follow-up workloads (US Census Bureau, 2017; p. 22). Administrative record use for these purposes has been identified as a high-level risk to the 2020 Census (Blumerman & Fontenot, 2017). Unlike online response, administrative record use may not be a census strategy of which many respondents are aware. However, previous research has shown that how administrative record use is framed has an impact on its favorability (Singer et al. 2011; Childs 2015; Childs, Walejko, and Eggleston 2015). People are also more skeptical of the Census Bureau's ability to keep data secure and confidential when sharing between agencies, which will occur when administrative records are used (Childs et al. 2015a). While some groups support the use of administrative records to replace or prepopulate census forms, a misstep in this area as we move toward a records-based 2030 Census could jeopardize the trust that the public has in the Census Bureau

(Mitre, 2017; JASON 2016). This evaluation allows us to collect up-to-date feedback from respondents on administrative record use in the census environment.

This evaluation also provides an opportunity to investigate any potential link between privacy and confidentiality concerns and mismatches between self-reported data and administrative records. It is an open question whether people with more privacy and confidentiality concerns are more likely to have mismatches between survey responses and administrative records, or missing administrative records. By linking responses to this evaluation with administrative records we could begin to address the relationship between privacy concerns and consistency of administrative records with self-reported data.

Concerns about privacy and confidentiality will continue to shape how the Census Bureau interacts with the public and how we address these concerns is of critical importance as we execute the 2020 Census and begin to prepare for 2030. There is growing, recent evidence that these types of concerns are increasingly salient and if unaddressed could contribute to the undercount of certain populations and item nonresponse (CSM 2017).

It is critical to conduct this research within the 2020 Census environment for three related reasons. First, the decennial census environment is unique in that the salience of government data collection will likely be quite high for most Americans. Assessing respondents' concerns with government data collection shortly after having made a decision about whether and how to share data with the government is an opportunity to gauge attitudes about privacy and confidentiality more accurately than at other times. Second, public discourse such as news media might also discuss matters of privacy and confidentiality during a decennial census that people might not often think about, helping to shape opinions and attitudes. Finally, the privacy and confidentiality concerns with the amount and types of data collected in a decennial census might also be different than those associated with a survey that has a smaller sample size but more in-depth data collection, such as the American Community Survey (ACS).

To fulfill our constitutional mandate, the 2020 and 2030 censuses will be used to apportion districts for representation in Congress. Public trust in the accuracy and reliability of the census will be important to support the fulfilling of that mandate, and this evaluation will provide us with the tools to craft messaging and approaches that will ensure that trust.

II.   Background

This work continues studies conducted as part of previous decennial census evaluations as well as more recent work that has been ongoing throughout this decade. Surveys of privacy and confidentiality concerns were undertaken as part of the 1990 and 2000 decennial census evaluation programs and this evaluation continues that work. This decade, the Gallup Daily Tracking Survey and Census Test Focus Groups provide background for this evaluation.

In a follow-up to the 1990 Census, the Census Bureau contracted with NORC to conduct a nationwide in-person survey focusing on issues related to census participation. One of the concerns that the survey was designed to address was privacy and confidentiality concerns with the data. The questionnaire included items about general privacy concerns and items specific to

the census. The sample was nationally representative and also included nonrespondents to the census. Data from this survey was linked to actual census response by asking respondents to provide an address for the purpose of matching back to census response records. This design allowed Singer et al. to analyze the role of privacy and confidentiality in census participation. The authors found that privacy and confidentiality concerns explained around 1.5 percent of the variance in return rates after controlling for demographics (Singer et al. 1993).

Similarly, after the 2000 Census, the University of Michigan, under contract with the Census Bureau, collected data with the Gallup organization to examine trends in beliefs about confidentiality and privacy. This study also investigated trends in attitudes toward data sharing. Again this evaluation matched back to actual census responses by asking respondents for their address for matching purposes. As in the 1990 Census evaluation, this study also found that privacy and confidentiality concerns explained about 1.5 percent of the variance in the mail return rate after controlling for demographics (Singer et al. 2003). This study also found increasing concerns about the sharing of confidential data among federal agencies.

In 2011, the Census Bureau's Communications Directorate conducted the second iteration of the Census Barriers, Attitudes, and Motivators Survey (CBAMS II) as a follow-up to the original CBAMS conducted prior to the 2010 Census in 2008. CBAMS was conducted to gain an in-depth understanding of the public's opinions about the 2010 Census, with the specific intention to understand those who have negative attitudes toward the census and the government more generally or those who are unaware/lack extensive knowledge of the census. CBAMS II provided a post-2010 Census measurement of the same issues as well as information on the use of administrative records for the decennial census. In CBAMS II, respondents were experimentally divided into three groups in order to test their views of administrative records use as a means of (1) reducing census (government) costs, (2) reducing respondent burden or (3) as simply an alternative option to a self-response (the control group). From this research, the study found that both arguments of reducing cost (when citing a $10 billion census price tag) and of alleviating respondent burden increased public support of administrative records usage, though the cost reduction frame was more powerful (Wroblewski, Bates and Pascale, 2012; Conrey, ZuWallack, and Locke, 2011).

Additionally, the CBAMS II found that some administrative records are less sensitive than others. People were more comfortable with obtaining one's name, date of birth, gender, and race from tax returns (50 percent), or other government records such as unemployment or social security (45 percent); whereas they were much less in favor of the census obtaining credit bureau data (25 percent) or medical records (22 percent) for use in a decennial census. Further, in the study, most people (65 percent) expressed unwillingness to allow the Census Bureau to use social security numbers to obtain sex, age, date of birth, and race information from other government agencies. Other research has suggested the importance of providing a context for answering such questions, and CBAMS II, like many telephone surveys, afforded limited opportunity to provide such context.

Beginning in February 2012, the Census Bureau has asked a random sample of approximately 200 respondents nightly questions on trust, confidentiality, credibility, transparency, and data use

as part of the Gallup Daily Tracking Survey. The data from this survey provides us with a time-series on trust in federal statistics that can be extended by this proposed research on privacy and confidentiality.

Between 2014 and 2016, the Center for Behavioral Science Methods has also conducted focus groups as follow-up research to annual census tests. The groups focused on privacy and confidentiality concerns of different segments of the populations in scope for each test. Groups included both respondents from different modes and nonrespondents and were separated demographically (e.g., by age, race, and language) when feasible. Findings from these focus groups suggest that demographics are indicative of important differences in terms of the types of privacy and confidentiality concerns that people have (Morales et al. 2017; Fobia et al. forthcoming). Related research also indicates that Spanish-language speakers have particular concerns about privacy and confidentiality as well (CSM 2017; Sha et. al 2018).

The Gallup Daily Tracking Survey and focus groups inform this work by suggesting the types of privacy and confidentiality concerns that might be prominent for census respondents. For example, in terms of responding to the census online, respondents in the 2014 Census Test focus groups were concerned about individuals posing as the Census Bureau via malicious links or contact attempts and stealing information (Holzberg & Fobia, 2016). In the Gallup survey, respondents who had concerns about answering the census online also report being concerned about hacking and data security (Childs et al., 2017). In terms of administrative record use, some respondents in 2014 and 2015 Census Test focus groups thought that government agencies should not share information with each other (Holzberg & Fobia, 2016; Morales, Holzberg, & Eggleston, 2017). Lack of trust in the government and concerns about hacking were reasons why some Gallup respondents did not support administrative records use (Childs et al. 2015a).

Meanwhile, other research has shown that rates of both consent to link data and overall survey participation have declined, raising concerns about the accuracy of results drawn from linked data and survey responses (Fulton 2012; Sakshaug and Kreuter 2012; Curtain, Presser, & Singer 2005; National Research Council 2013). A study by Singer and Presser (1996) demonstrated that individuals' reactions to data-sharing arrangements (to facilitate mandatory census activities) were influenced by demographics, especially gender and education. Research by Huang, Shih, Chang, and Chou (2007) in Taiwan revealed that the elderly, lower income, less educated, and minorities were less likely to consent to sharing and linking their information for research purposes, but that gender was not a factor.

In addition to demographics, some of respondents' opinions and knowledge are also related to one's openness to data linkage. For example, research by Singer and Presser (1996) established that people's propensity to share or link their data was swayed by their understanding of the statistical agencies involved in those endeavors, their belief that the information is already being shared, and the importance they attach to the use of shared information. Similarly, negative attitudes toward the use of administrative records have also attributed to respondents' lack of understanding of what administrative records are, how statistical agencies make use of that information, the authority of the statistical agencies, and their ability to protect confidentiality (Bates and Pan, 2009; Gerber and Landreth, 2007, Holzberg & Fobia, 2016; Morales, Holzberg, & Eggleston, 2017).

Respondents' reactions may also not always be in the direction we might expect, and therefore they should continue to be studied. For example, previous research has shown respondents to be less favorable to the use of administrative records to determine the occupancy of housing units than they are for administrative record use to fill in basic census demographic information (Childs et al. 2015a).

Privacy and confidentiality concerns have been cited as a potential reason for nonresponse in web surveys in particular (Couper 2000; Cho and LaRose 1999). In the 2020 Census there will be three modes of self-response available and an in-person Nonresponse Followup (NRFU). In this evaluation, we will investigate how privacy and confidentiality concerns might affect self-response choices between online, mail and telephone response as well as a respondent being enumerated in-person in the 2020 Census. Another behavior we will study is the relationship between distinct privacy and confidentiality concerns and item nonresponse. Literature suggests that item nonresponse is connected to respondent concerns with confidentiality of disclosure (Booth-Kewley et al.,2007; Joinson et al., 2004). Concerns about privacy and confidentiality might be related to specific items asked on the decennial census questionnaire that might be seen as sensitive.

The relationship between respondents' privacy and confidentiality concerns and the likelihood that their self-response data does not match administrative records is a gap in the literature on discrepancies between records and survey data that we plan to address. We plan to link responses to our evaluation survey with administrative record data. If increased privacy and confidentiality concerns are related to a higher likelihood of a mismatch between self-reported and administrative records data it could indicate bias in either the records or self-reported data. This evaluation could provide a starting point for further research into record mismatches as part of the 2030 Census research program.

This evaluation, in line with evaluations in earlier decades, will connect privacy and confidentiality concerns with respondent behavior in a decennial census environment. What our research from this decade has shown is that respondents have particular concerns about responding to surveys online. However, much of that data collection has been hypothetical and qualitative. We will investigate the relationship of privacy and confidentiality concerns to response mode, item nonresponse, and mismatches between administrative records and self-reported data.

III.   Assumptions

1. The project team will obtain adequate funding to implement the evaluation as it is designed in this study plan.
2. The 2020 Census will have an online response option.
3. The 2020 Census will use administrative records for operations as planned.
4.   The Census Data Lake will contain 2020 Census response and operational data required for analysis.

5. The project team assumes that the Census Bureau will be able to obtain the services of a contractor to support the design and implementation of this evaluation.

IV.   Research Questions

Three research questions are central to this project:

1. Are privacy and confidentiality concerns related to response mode?
   a. How do these concerns vary by demographic group?

In the 2020 Census there will be three modes of self-response available and an in-person follow-up. In this evaluation, we will investigate how privacy and confidentiality concerns might affect self-response choices between online, mail, and telephone response. Privacy and confidentiality concerns might also be related to a respondent being enumerated in-person rather than self-responding.

2. Are privacy and confidentiality concerns related to partial responses?
   a. How does this relationship vary by demographic group?

Based on data completeness measures from the 2010 Census, we expect 89 percent of self-response forms to include all five person-level variables while item nonresponse rates for household-level items range from 1.8 percent (household count) to 7.8 percent (telephone number) (Rothhaas et al. 2012). For this project, we define a partial response as missing one or more items. We expect nonresponse rates to different items to be related to privacy and confidentiality concerns.

3. Are privacy and confidentiality concerns related to mismatches between administrative records and self-reported data?
   a. How does this relationship vary by demographic group?

We plan to link survey responses from this evaluation with Internal Revenue Service (IRS) and/or Social Security Administration (SSA) records. We are using decennial response data to select our sample so decennial responses will be available. A mismatch occurs when the self-response data from the decennial responses or from the evaluation survey response do not match data from IRS or SSA records. We expect that households with mismatches will have different or increased privacy and confidentiality concerns when compared with households that do not have mismatches.

V.   Methodology

In this section, we detail the methodology for a survey of decennial census respondents' privacy and confidentiality concerns. Data collection will begin shortly after April 1, 2020, and include both telephone and in-person modes. The survey will be administered in both English and Spanish. The sample will focus on detecting differences in demographic groups. The instrument will take between 20 and 35 minutes to administer and include question items in four topic areas:

1) privacy and confidentiality concerns, 2) opinions on administrative records, 3) concerns about decennial census items, and 4) related constructs (see page 11 for details). Analysis plans include logistic regression models, multinomial logistic regression, and t-tests for differences between demographic groups. Our analysis plan also includes linking survey responses with administrative records from the IRS and/or SSA.

This study plan also includes a qualitative component. Some respondents with privacy and confidentiality concerns will likely not complete the census or allow entrance to enumerators and observers. To capture these respondents we will conduct a qualitative study that leverages the 2020 Census Partnership Program in addition to the survey. The qualitative study will include three components: 1) observations of events, 2) interviews with national partners, and 3) focus groups with community partners.

A. Evaluation Design

Data Collection

Quantitative Component

Data will be collected in two modes: telephone and in-person. Self-respondents to the decennial census will be interviewed by telephone and NRFU census respondents will be interviewed in-person and may be offered a small incentive. The survey will be conducted in both English and Spanish.

Telephone Survey. Telephone data collection will use an instrument programmed by a contractor. Telephone interviewers would make the contact attempts to ask respondents to participate and administer the instrument. 2020 Census respondent-provided phone numbers would be used to contact respondents. Contact frame telephone numbers may be used if no respondent-provided phone numbers have been collected, or if the reuse of respondent-provided phone numbers is prohibited for this purpose. Then, a sample for the telephone survey would be drawn from census response data on a flow basis and sent to telephone interviewers for follow-up. We would like the follow-up survey to be conducted in close proximity to when a respondent fills out their census form. Ideally, data collection would begin in April.

In-Person Survey. Census respondents who fill out their forms with enumerators during NRFU will be sampled for an in-person follow-up survey. A sample of addresses that do not respond to the telephone survey will also be selected for in-person follow-up. The interviewers that would be used for this task will be employed by the contractor.

Sample

The universe for this evaluation encompasses households that responded to the decennial census (omitting households selected for other decennial census experiments and evaluations when necessary), the ACS, and the census Post-Enumeration Survey. We will draw sample using 2020 Census response data for Person 1. Based on findings from previous research, we are primarily interested in three race/ethnicity groups: Hispanic (any race), White (alone, non-Hispanic), and Black (alone, non-Hispanic).[1] We plan to sample so that we will be able to cross these race groups with age groups (18-24, 25-44, 45-64, 65+). We will include both self-respondents and NRFU respondents in our sample.

Using an alpha of 0.10, beta of 0.20, and a detectable difference of 8 percentage points in privacy concerns, the national sample size necessary for this evaluation is 103,340 housing units. We will draw sample during the 2020 Census data collection using characteristics of the 2020 Census return (e.g., response mode, complete or partial response, race data, age data).

Self-responses to the 2020 Census with available phone numbers, either provided by the respondent or from the Census Bureau's contact frame, are first sorted by geography, partial response, 2020 Census response mode, contact strategy, and language. Next, these are stratified by the race and age stratums of interest: non-Hispanic White ages 18-29, non-Hispanic White ages 30-44, non-Hispanic White ages 45-59, non-Hispanic White ages 60+, non-Hispanic Black ages 18-29, non-Hispanic Black ages 30-44, non-Hispanic Black ages 45-59, non-Hispanic Black ages 60+, Hispanic ages 18-29, Hispanic ages 30-44, and Hispanic ages 45-59. Finally, a systematic random sample is taken to obtain a sample of 9,059 housing units from each of the race and age stratums of interest resulting in a total of 99,653 housing units selected from self-responses.

The NRFU responses are sorted by geography, contact strategy, by the race and age stratums of interest: non-Hispanic White ages 18-29, non-Hispanic White ages 30-44, non-Hispanic White ages 45-59, non-Hispanic White ages 60+, non-Hispanic Black ages 18-29, non-Hispanic Black ages 30-44, non-Hispanic Black ages 45-59, non-Hispanic Black ages 60+, Hispanic ages 18-29, Hispanic ages 30-44, and Hispanic ages 45-59, and language before taking a systematic random sample to obtain the 3,687 NRFU housing units.

Instrument

The instrument assesses respondents' privacy and confidentiality concerns. We will ask questions that surround four themes: 1) privacy and confidentiality concerns, 2) opinions on the use of administrative records, 3) concerns about particular census questions, and 4) other related constructs.

---

[1] Previous research has shown differences in privacy and confidentiality attitudes by race, Hispanic origin, sex, and age. See Morales et al. 2017; Fobia et al. forthcoming; Sha et al. 2018; CSM 2017.

The instrument will be written by Center for Behavioral Science Methods (CBSM) staff and cognitively tested in English and Spanish. We anticipate that the instrument will take between 20 and 35 minutes to administer.

Privacy and Confidentiality Concerns. Respondents will be asked if they have any privacy or confidentiality concerns about their census responses and data; interviewers will code the type of concern and ask follow up questions about the level of respondent concern. Additionally, respondents will be asked to choose a level of concern for different types of privacy and confidentiality concerns that have been found in previous research or identified as emerging issues. Some of these may include hacking, misuse of data, the government having too much information, re-identification, data sharing, and computer scams. Respondents will be asked about their perceptions of each of the four Census Bureau privacy principles: necessity, openness, respect for respondents, and confidentiality (U.S. Census Bureau, 2006). The privacy and confidentiality practices asked about will also include the differential privacy methods that the Census Bureau plans to implement. This project is an opportunity to evaluate respondent confidence and beliefs about our practices in this arena.

We will replicate questions from privacy and confidentiality studies in past decades. In particular, we plan to replicate questions about privacy beliefs, confidentiality concerns, and opinions about administrative records (Singer et al. 1993; Singer et al. 2003).

Opinions on Administrative Records. Results from attitude questions about administrative records will inform the research and testing phase of a records-based 2030 Census. The items for this topic will be replicated from the Gallup Nightly Survey as well as previous decennial evaluations of privacy and confidentiality (See Singer et al. 1993; Singer et al. 2003; CBAMS II Final Report). We plan to ask respondents for their income information. Past research has shown that failure to report income on a public opinion survey is highly correlated with reported privacy and confidentiality concerns.

Concerns about decennial census items. We will also ask respondents about their level of privacy and confidentiality concern for each of the census items. This data will help us understand whether certain items are more sensitive than others and whether the sensitivity of particular items are associated with different demographic groups. This evaluation will provide data about privacy and confidentiality concerns about the citizenship question that is planned for the 2020 Census questionnaire. Data on individual census items can also help inform decision making around privacy budgets for data releases that use differential privacy methods.

Related constructs. Other questions will replicate those asked in the Gallup Nightly Survey that have been shown in previous studies to be related to privacy and confidentiality concerns. For

example, respondents may be asked questions about their knowledge of federal statistics as well as trust in the federal government, the Census Bureau, and other institutions. In previous work, knowledge about the federal statistical system, data use, and belief in the relevance of statistics are important correlates of trust in federal statistics and self-reported response (Childs et al. 2015; Childs et al. 2017; Conrey et. al. 2012). Trust in government has also been identified as an important challenge for the 2030 Census (Mitre, 2016).

Analysis

We will begin by running correlations between our predictor and outcome variables (See Appendix A, Table x1). We may create indices for privacy and confidentiality concerns depending on the final items selected for the questionnaire. We will run descriptive statistics on attitude items and outcome variables as well as other exploratory analyses in addition to what is described below.

1. Are privacy and confidentiality concerns related to response mode?
    a. How do these concerns vary by demographic group?

For the first research question, we will use a multinomial logistic regression model to test the relationship between response mode and our predictor variables. For this model, the predictor variables include items about privacy and confidentiality concerns, related constructs, concerns about census items, demographics, and whether or not the respondent reported income (See Appendix A, Table RQ1). Base models will not include demographic controls.

2. Are privacy and confidentiality concerns related to partial responses?
    a. How does this relationship vary by demographic group?

For the second question, we will use logistic regression models to test the relationship between partial response and our predictor variables. We will have a binary outcome variable for complete versus partial response. We will also run models for binary outcome variables for each missing data item (e.g. missing citizenship versus not missing citizenship, missing birthdate versus not missing birthdate, etc.). The predictor variables include items about privacy and confidentiality concerns, related constructs, concerns about census items, demographics and whether or not the respondent reported income (See Appendix A, Table RQ2). Base models will not include demographic controls.

3. Are privacy and confidentiality concerns related to mismatches between administrative records and self-reported data?
    a. How does this relationship vary by demographic group?

For the third research question, we will link responses from this evaluation survey to IRS and/or SSA administrative records. We will use logistic regression models to test the relationship between administrative records and self-reported data mismatches. Decennial response data will provide the self-reported items to be compared to IRS and SSA data on the same items. We will ask for respondent income in the evaluation survey. We will create binary outcome variables for

discrepancies for each data item (e.g. number of people in household reported in the 2020 Census does not match most current administrative records versus data items match correctly). Predictor variables include items about privacy and confidentiality concerns, opinions on administrative records, related constructs, concerns about census items, and demographics (See Appendix A, Table RQ3).

For all three subquestions, we will use a chi-square test to compare the distributions of responses by demographic groups. If significant differences are found ($p < 0.10$), we will run t-tests adjusted for multiple comparisons using a Bonferroni adjustment to further examine the pattern of these differences.

Qualitative Component

Since it is likely that some respondents with privacy and confidentiality concerns will not complete the census or allow entrance to enumerators and observers, we will conduct a qualitative study that leverages the 2020 Census Partnership Program in addition to the survey. The qualitative study will include four main components: 1) Qualitative interviews with cultural experts recommended by the partnership program, 2) focus groups with trusted messengers, 3) observations of partnership events, and 4) focus groups with respondents.

This component complements research on the effects of the citizenship question on respondent participation as well as on the survey of privacy and confidentiality concerns. This qualitative component is not representative research and the findings will be limited in their generalizability to larger populations. However, since people who do not respond to the 2020 Census are not likely to be captured using other methods, this aspect of the research will fill this gap.

Research Questions

1. What can community partners tell us about reasons people in their communities do not complete census forms?
2. What effect did including the citizenship question on the 2020 Census questionnaire have on participation? What are the reasons respondents and community partners give for this effect?
3. What other privacy and confidentiality concerns are expressed, if any?

Methodology

The study will include four main components: 1) Qualitative interviews with cultural experts recommended by the partnership program, 2) focus groups with trusted messengers, 3) observations of partnership events, and 4) focus groups with respondents.

We will leverage the partnership program for this study since people who might have concerns about the citizenship question will likely not speak directly with government employees or contractors. The partnership program seeks to partner with people in hard-to-count (HTC) communities who are already trusted in those communities.

Qualitative interviews with cultural experts. Cultural experts are individuals who will be recommended by the 2020 National Partnership Program (NPP). These individuals will come from organizations that have experience working with HTC communities and knowledge of and access to networks of trusted messengers. In depth, qualitative interviews with cultural experts will allow us to draw on the experience of those that have been able to successfully reach HTC groups and have chosen to be part of the Partnership Program. Partnership specialists will play a key role by connecting researchers to partners. These interviews will identify concerns they have encountered about respondent participation in the 2020 Census because of the citizenship question and will collect information about strategies they used to increase response and their effectiveness. Interviews will be both in-person and by telephone when necessary. We plan to conduct 15 interviews in 2020.

Focus groups with trusted messengers. Trusted messengers are individuals who are influential in their local communities and may affect others' decisions about when, how, and whether to respond to the census or other surveys. These are also individuals who are targeted through the 2020 Community Partnership and Engagement Program (CPEP). In the context of the decennial census, CPEP aims to engage community partners to increase decennial participation of those who are less likely to respond or are often missed. While the National Partnership Program partners with larger organizations, the 2020 CPEP will engage at the grassroots level to reach out to those who are less likely to respond to the national campaign (Hall 2017). CPEP plans to leverage trusted messengers (also called "trusted voices") to increase response rates in hard-to-count populations. Our research goal with this group is to learn more about concerns trusted messengers may have about respondent participation in the 2020 Census, how they addressed these concerns with respondents, and whether their strategies were effective. We plan to conduct six focus groups in 2020.

Observations of partnership events. Researchers will observe 2020 Partnership Program events at both the national level and community program level. One example of an event is the Census Solutions Workshops hosted by members of the NPP. By observing these types of events, we can see how communities are engaging with the 2020 Census, the questions and concerns that respondents are mentioning, and how the partners respond to those questions and concerns. From these preliminary interviews, we will ask respondents about planned events or examples of grassroots efforts that we can observe in the next phase of the research. We plan to observe 10 events: five at the NPP level and five at the CPEP level.

Focus groups with respondents. Researchers will conduct focus groups with members of hard-to-count populations who might have been impacted by the inclusion of the citizenship questions. Focus groups will be conducted in English and other languages. This set of focus groups will allow us to assess the impact of privacy and confidentiality concerns for respondents who might not be captured in the larger quantitative survey because of the sampling strategy and sample size constraints. We plan to conduct about 12 focus groups with respondents across the United States.

Researchers could be Census Bureau staff or contractors. Some researchers for certain populations will need to be bilingual English-Spanish speakers. Other languages might also be necessary.

Analysis

Researchers will complete individual summaries for each qualitative interview and observation event. Focus groups will be transcribed and translated into English when necessary. Summaries and transcripts will be reviewed for evidence of recurring themes and patterns. Each analyst will code the interview summaries, observation summaries, and focus group transcripts separately before meeting as a group. The research team will then discuss the recurring themes and patterns in the data and work to reach a consensus on the themes and codes for reporting and conduct further analysis if needed.

B. Interventions with the 2020 Census

This proposal does not require direct interventions with the 2020 Census systems or processes. However, it does propose to use 2020 systems independently from production. We will need response data to draw sample and to make sure that all cases are unduplicated with other decennial census experiments and evaluations cases when necessary and ACS sampled households.

The response data to the follow-up should probably ultimately reside in the Census Data Lake (CDL). That means there needs to be a connection somewhere into the CDL, and CDL has to expect the file.

C. Implications for 2030 Census Design Decisions and Future Research and Testing

As we begin the research and testing phase for the 2030 Census, this evaluation will provide a starting point for research on respondents' perceptions and understanding of how the census uses administrative records. Future research into how to communicate around a primarily records-based census could build on the results of this study about the relationship of privacy and confidentiality concerns and mismatches between administrative records and self-reported data. This evaluation is also designed to detect differences in privacy and confidentiality concerns for different demographic groups. This data can inform the design of qualitative studies to take place throughout the next decade to analyze the meanings that respondents attach to the census and other surveys and how that will change with a records-based census. It will also provide a baseline for other quantitative studies to assess changes in privacy and confidentiality concerns as the coming decade progresses. The data from this evaluation will also inform the way that the Census Bureau communicates privacy protections with its respondents in the coming decade. We will gather nationally representative data on the relationship between item nonresponse and concerns about privacy and confidentiality. This can inform messaging as well as decisions around privacy budgets for specific data points for this and other surveys.

The results will potentially lead to cost savings by informing our messaging on privacy and confidentiality, which can possibly lead to increased unit response in the 2030 Census. The

proposed research will also provide some insight into how to reduce item nonresponse, as some respondents may choose to skip questions because of privacy and confidentiality concerns. This evaluation will collect information that can help inform messaging for the Census Bureau to use to reassure members of the public and stakeholders who are concerned about privacy and confidentiality. Privacy and confidentiality concerns have a longstanding history in the census and Census Bureau surveys. For example, the contact centers and respondent advocate field phone calls and emails from respondents who find the ACS to be "too intrusive" or who are unsure whether it is a legitimate request or a scam. These concerns will likely increase as technology evolves throughout the next decade. Research in this area and related areas such as differential privacy and data disclosure avoidance is of critical importance.

VI.   Data Requirements

Data File/Report	Source	Purpose	Expected Delivery Date
Census response file	Census Data Lake	Universe for sample	mm/dd/yyyy
Decennial response event file	CDL	Universe for sample	mm/dd/yyyy
Contact frame phone numbers	CARRA/PEARSIS	Append phone numbers to sample cases	mm/dd/yyyy

VII. Risks

There are no risks that impact the completion of this evaluation.

VIII. Limitations

1. We do not currently plan to survey people who did not respond to the decennial census. This would significantly increase the costs of this evaluation, but would also increase the value of the results.

IX.   Issues That Need to be Resolved

1. It is possible that there may be some respondents who find issues of privacy and confidentiality sensitive in this evaluation. We will carefully script a survey introduction to reassure these respondents to the maximum extent possible.
2. It is an open question whether we can link the data from this survey with administrative records from IRS and/or SSA.

X.   Division Responsibilities

Division or Office	Responsibilities
CBSM	• Project design and coordination • Drafting and pretesting of instrument • Analysis • Design of qualitative component • Qualitative data collection and analysis
DSSD	• Project design • Sample design and specifications
Contractor	• Telephone and in-person instrument and case management • Telephone and in-person data collection • In-person incentive management • Qualitative data collection and analysis

XI.   Milestone Schedule

Privacy and Confidentiality Evaluation Milestone	Date
Begin qualitative data collection	03/2020
Draw universe from Census Data Lake	04/2020-09/2020
Select sample for follow ups	04/2020-09/2020
Begin telephone data collection	04/2020
Begin in-person data collection	06/2020
Wrap telephone data collection; send sample to in-person follow-up	08/2020
Wrap in-person data collection	09/2020
Wrap qualitative data collection	09/2020
Receive, Verify, and Validate Data For Privacy and Confidentiality Evaluation	mm/dd/yyyy
Distribute Initial Draft Privacy and Confidentiality Evaluation Report to the Decennial Research Objectives and Methods (DROM) Working Group for Pre-Briefing Review	mm/dd/yyyy
Decennial Census Communications Office (DCCO) Staff Formally Release the FINAL Privacy and Confidentiality Evaluation Report in the 2020 Memorandum Series	mm/dd/yyyy

XII. Review/Approval Table

Role	Approval Date
Primary Author's Division Chief (or designee)	08/21/2018
Decennial Census Management Division (DCMD) ADC for Nonresponse, Evaluations, and Experiments	02/19/2019
Decennial Research Objectives and Methods (DROM) Working Group	02/19/2019
Decennial Census Communications Office (DCCO)	mm/dd/yyyy

XIII. Document Revision and Version Control History

Version/Editor	Date	Revision Description
1.0/Fobia	8/10/2018	Original
2.0/Fobia	2/07/2019	Revised after DROM 10/2/2018
3.0/Fobia	3/15/2019	Revised after DROM review 2/19/2019

XIV. Glossary of Acronyms

Acronym	Definition
ACS	American Community Survey
ADC	Assistant Division Chief
CBAMS	Census Barriers, Attitudes, and Motivators Survey
CBSM	Center for Behavioral Science Methods
CDL	Census Data Lake
CPEP	Community Partnership and Engagement Program
CSM	Center for Survey Measurement
DCCO	Decennial Census Communications Office
DCMD	Decennial Content Management Division
DROM	Decennial Research Objectives and Methods Working Group
DSSD	Decennial Statistical Studies Division
EXC	Evaluations & Experiments Coordination Branch
HTC	Hard-to-Count
IPT	Integrated Project Team
IRS	Internal Revenue Service
NPP	National Partnership Program
NRFU	Nonresponse Followup
R&M	Research & Methodology Directorate
SSA	Social Security Administration

XV.   References

Bates, N.A. and Yuling, P. (2009). "Motivating Non-English Speaking Populations for Census and Survey Participation" Paper presented at the Federal Committee on Statistical Methodology 2009 Research Conference, Washington DC, November 3, 2009.

Bates, N., Wroblewski, M., & Pascale, J. (2012). Public attitudes toward the use of administrative records in the U.S. Census: Does question frame matter? Technical Report, *Survey Methodology Series. #2012-04*, United States Bureau. Available at: https://www.census.gov/srd/CBSMreports/byrsm.html#y2012.

Blumerman, L., & Fontenot, A. (2017). 2020 Census Program Management Review: Welcome and High-Level Program Updates. Washington, DC: U.S. Census Bureau. Retrieved from: https://www2.census.gov/programs-surveys/decennial/2020/program-management/pmr-materials/07-11-2017/pmr-welcome-high-level-updates-07-11-2017.pdf

Booth-Kewley, S. Larson, G. & Miyoshi, D. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior.* Vol 23:1. https://doi.org/10.1016/j.chb.2004.10.020

Center for Behavioral Science Methods (CBSM). (2017). Respondent Confidentiality Concerns, Memorandum for ADRM.

Childs, J.H., Eggleston, C., Morales, G., & Fobia, A.C. (2017). Attitudes Towards Relevance of Statistics in Public Policy-Making. Paper presented at the meeting of the World Association for Public Opinion Research, Lisbon, Portugal.

Childs, J.H., King, R. & Fobia A.C. (2015). Confidence in US Federal Statistical Agencies. Survey Practice. Volume 8. Issue 5.

Childs, J.H., King, R., Eggleston, C., & Holzberg, J. (2015a). "Public Attitudes towards Use of Administrative Records to Supplement the U.S. Census." Paper presented at the 68th Annual Conference of the World Association for Public Opinion Research, Buenos Aires, Argentina.

Cho, H., & Larose, R. (1999). "Privacy Issues in Internet Surveys" *Social Science Computer Review.* Vol. 17: 4. https://doi.org/10.1177/089443939901700402

Conrey, F. R., ZuWallack R. & Locke, R. (2012). Census Barriers, Attitudes, and Motivators Survey II Final Report. Washington, DC: U.S. Census Bureau. Retrieved from: https://www.census.gov/2010census/pdf/2010_Census_CBAMS_II.pdf

Couper, M. (2000). Review: Web Surveys: A Review of Issues and Approaches. The Public Opinion Quarterly, 64(4), 464-494. Retrieved from http://www.jstor.org/stable/3078739

Curtin, Richard, Stanley Presser, and Eleanor Singer. 2005. Changes in Telephone Survey Nonresponse over the Past Quarter Century, Public Opinion Quarterly, Volume 69, Issue 1, 1 January 2005, Pages 87–98, https://doi.org/10.1093/poq/nfi002.

Das, Marcel, and Mick P. Couper. 2014. "Optimizing Opt-Out Consent for Record Linkage." *Journal of Official Statistics* 30(3):479-497.

Gerber, E. and Landreth, A. (2007). "Respondents' Understandings of Confidentiality in a Changing Privacy Environment," *U.S. Census Bureau, Statistical Research Division.*

Huang, N., Shih, S., Chang, H., and Chou, Y. (2007). "Record Linkage Research and Informed Consent: Who Consents?" BMC Health Services Research, 7:18 (February 2007).
Fobia, A. C., Morales, G. and Childs, J.C. (forthcoming). 2020 Census Research and Testing: Final Report on the 2016 Census Test Focus Groups. Washington, DC: U.S. Census Bureau.

Fulton, Jenna. 2012. "Respondent Consent to Use Administrative Data." Doctoral dissertation, University of Maryland.

Holzberg, J.L., & Fobia, A.C. (2016). 2020 Census Research and Testing: Final Report on the 2014 Census Test Focus Groups (2020 Census Program Internal Memorandum Series: 2016.43.i). Washington, DC: U.S. Census Bureau.

JASON Program Office, Mitre Corporation (2016). Alternative Futures for the Conduct of the 2030 Census.

Joinson, A., Woodley, A., & Reips, U. (2004). Personalization, authentication, and self-disclosure in self-administered Internet surveys. *Computers in Human Behavior,* 23, 275–285. doi:10.1016/j.chb.2004.10.012.

Mitre Corporation. (2017). US Census and American Enterprise Institute: Census 2030 Blue Sky Event.

Mitre Corporation. (2016). 2030 and the Census Environment. A Discussion Sponsored by the Brookings Institution Hutchins Center on Fiscal and Monetary Policy.

Morales, G., Holzberg, J., & Eggleston, C. (2017). 2015 Census Test Focus Groups: Final Report (2020 Census Program Internal Memorandum Series: 2017.4.i). Washington, DC: U.S. Census Bureau.

National Research Council. 2013. *Nonresponse in Social Science Surveys: A Research Agenda.* R. Tourangeau and T.J. Plewes (Eds.) Panel on a Research Agenda for the future of Social Science Data Collection, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Pascale, Joanne. 2011. "Requesting Consent to Link Survey Data to Administrative Records: Results from a Split-Ballot Experiment in the Survey of Health Insurance and Program Benefit

Wording and Data-Linkage Consent." Technical Report, Survey Methodology Series #2011-03, United States Census Bureau.

Rothaas, C., F. Lestina and J. Hill. (2012). 2010 Decennial Census: Item Nonresponse and Imputation Assessment Report. Decennial Statistical Studies Division. Washington, DC: U.S. Census Bureau.

Sakshaug, Joseph W., and Frauke Kreuter. 2014. "The Effect of Benefit Wording on Consent to Link Survey and Administrative Records in a Web Survey." *Public Opinion Quarterly* 78(1):166-176.

Sakshaug, Joseph W., Mick P. Couper, Mary Beth Ofstedal, and David. R. Weir. 2012. "Linking Survey and Administrative Records: Mechanisms of Consent." *Sociological Methods & Research* 41:535–69.

Sakshaug, Joseph, Valerie Tutz, and Frauke Kreuter. 2013. "Placement, Wording, and Interviewers: Identifying Correlates of Consent to Link Survey and Administrative Data." *Survey Research Methods* 7:133–144.

Sha, M., Son, J., Pan, Y., Park, H., Schoua-Glusberg, A., Tasfaye, C., Sandoval Giron, A., García Trejo, A., Terry, R., Goerman, P., Meyers, M., and L. Lykke. 2018. "Multilingual Research for Interviewer Doorstep Messages Final Report." Research Report Series. 2018.

Singer, Eleanor, Nancy Bates, and John Van Hoewyk. 2011. *"Concerns about Privacy, Trust in Government, and Willingness to Use Administrative Records to Improve the Decennial Census."* Paper presented at the 66th Annual Conference of the American Association for Public Opinion Research, Phoenix, AZ, USA. Available at: http://ww2.amstat.org/sections/SRMS/Proceedings/y2011/Files/400168.pdf.

Singer, E. and Presser, S. (1996). "Public Attitudes toward Data Sharing by Federal Agencies." Paper presented at the Census Bureau Annual Research Conference, Washington DC.

Singer, E. and Van Hoewyk, J. (2001). "Trends in Attitudes Toward Privacy, Confidentiality, Data Sharing, 1995-2000." Paper presented at the Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001.

Singer, E., Van Hoewyk, J., and Neugebauer, R.J. (2003). "Attitudes and Behavior: The Impact of Privacy and Confidentiality Concerns on Participation in the 2000 Census." Public Opinion Quarterly, 67, 368-384.

Singer, E, Van Thurn, D.R., and Miller, E.R. (1995) "Confidentiality Assurances and Response: A Quantitative Review of the Experimental Literature." The Public Opinion Quarterly, 59:1, 66.77.

Singer, E., Van Hoewyk, J., and Presser, S. (1997). "Public Attitudes Toward Data Sharing by Federal Agencies." In Record Linkage Techniques-1997: Proceedings of an International Workshop and Exposition (pp. 237-247). Washington, DC: National Academy Press.

Smirnova, M., & Scanlon, P. (2017). Experience and Cultural–Repertoire Based Avenues of Trust: An Analysis of Public Trust in Statistical Agencies and their Data. Social Policy and Society 16(2), pp. 219-236. doi: 10.1017/S147474641500069X.

U.S. Census Bureau. (2017). 2020 Census Operational Plan: A New Design for the 21st Century.

U.S. Census Bureau, (2016). American Community Survey, July 2016 Panel.

U.S. Census Bureau. (2006). U.S. Census Bureau Privacy Principles. Retrieved from: https://intranet.ecm.census.gov/apps/policyportal/Pages/PolicyViewer.aspx?docID=375.

Wroblewski, M., Bates, N., and Pascale, J. (2012). Public Attitudes toward the Use of Administrative Records in the U.S. Census: Does Question Frame Matter? Federal Committee on Statistical Methodology Research Conference. Washington, DC.

Appendix A: Planned Tables for 2020 Privacy and Confidentiality Evaluation

Table x1. Bivariate Correlations of Predictors and Outcome Variables

	Mode of response	Partial Response	Mismatch with admin recs
Concerns about privacy			
Attitudes towards administrative records			
Concerns about census items			
Demographics			
Race			
Age			
Race*Age			
Education			
Region			
Sex			
Marital status			
Employment status			
Income not reported			

Table RQ1. Demographic and attitudinal predictors of web response mode v. all other modes

	Odds Ratios	Regression coefficients
Privacy and confidentiality concerns		
Related constructs		
Concerns about census items		
Demographics		
Race		
Age		
Race*Age		
Education		
Region		
Sex		
Marital status		
Employment status		
Income not reported		

Table RQ2. Demographic and attitudinal predictors of complete vs. partial response

	Any missing data	Missing citizenship status	Missing birthdate
Privacy and confidentiality concerns			
Concerns about census items			
Related constructs			
Demographics			
Race			
Age			
Race*Age			
Education			
Region			
Sex			
Marital status			
Employment status			
Income not reported			

Table RQ3. Demographic and attitudinal predictors of mismatches between administrative records and self-reported data items

	Any mismatch	Household count mismatch	Income mismatch	Citizenship mismatch
Concerns about privacy				
Opinions on administrative records				
Related constructs				
Concerns about census items				
Demographics				
Race				
Age				
Race*Age				
Education				
Region				
Sex				
Marital status				
Employment status				
Income not reported				

Table x2. Privacy and confidentiality attitudes by demographic group			Concerns about privacy	Attitudes towards admin records	Concerns about census items	Related constructs	Mismatch admin records
Demographics							
Age	18-29						
	30-44						
	45-59						
	60+						
Race	White (alone), non-Hispanic						
	Black (alone), non-Hispanic						
	Hispanic						
	More than one race						
Race*Age	White (alone), non-Hispanic	18-29					
		30-44					
		45-59					
		60+					
	Black (alone), non-Hispanic	18-29					
		30-44					
		45-59					
		60+					
	Hispanic	18-29					
		30-44					
		45-59					
		60+					
	More than one race	18-29					
		30-44					
		45-59					
		60+					
Sex	Male						
	Female						
Education	Less than HS and HS/GED						
	Some college/associates						
	Bachelors						
	Post-Bachelor's						
Region							
Marital Status							
Employment Status							
Income not reported							