

Estimating the Effect of Asking About Citizenship on the U.S. Census

Results from a Randomized Controlled Trial

A Shorenstein Center Discussion Paper

March 2019

Updated April 2019



HARVARDKennedy School

SHORENSTEIN CENTER
on Media, Politics and Public Policy

The views expressed in Shorenstein Center Discussion Papers are those of the author(s) and do not necessarily reflect those of Harvard Kennedy School or of Harvard University.

Discussion Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important issues and challenges in media, politics and public policy. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

Authors

Matthew A. Baum, Kalb Professor of Global Communication, Harvard Kennedy School

Bryce J. Dietrich, Research Fellow, Shorenstein Center on Media, Politics, and Public Policy, Harvard Kennedy School

Rebecca Goldstein, PhD Candidate, Harvard University Department of Government

Maya Sen, Associate Professor of Public Policy, Harvard Kennedy School

Authors' names listed in alphabetical order.

Abstract and Introduction

The 2020 U.S. Census will, for the first time since 1950, ask about residents' citizenship status. The effect of doing so on census completion across different racial/ethnic groups is, however, unknown. We introduce the notion of contextual sensitivity to explain how seemingly innocuous questions can become costly to answer in certain political environments. Using this concept and a large survey experiment (n = 9,035 respondents), designed to mirror the appearance and substance of the 2020 Census, we find that asking about citizenship status significantly increases the percent of questions skipped, with particularly strong effects among Hispanics, and makes respondents less likely to report having members of their household who are of Hispanic ethnicity. When extrapolated to the general population, our results imply that asking about citizenship will reduce the number of Hispanics reported in the 2010 Census by approximately 6.07 million, or around 12.03 percent of the 2010 Hispanic population.

Whether it is defining Native Americans as non-citizens in 1800 or introducing a “mulatto” category in 1850, the classification of race and ethnicity on the U.S. Census has long been inherently political (Nobles 2000). This is why many paused when the Census Bureau announced it would include, for the first time since 1950, a question on residents' citizenship status on the 2020 Census. An obvious concern is that some residents may refuse to participate altogether. Another less-well understood concern is that such a question may make any omissions more difficult to interpret. Are respondents who fail to report a Hispanic household member doing so in order to avoid potential prosecution? Or are they simply forgetting to include pertinent information? This ultimately affects data quality, which carries broader implications for the way federal funds are allocated and congressional districts are apportioned.

Moreover, given the current conservative rhetoric and policies regarding immigrants and immigra-

tion, some U.S. residents – particularly Hispanics – may be primed to be skeptical of the government’s motives if the citizenship question is introduced. We argue this kind of surrounding political environment can make some, intrinsically innocuous questions more costly to answer. We refer to this as *contextual sensitivity*. For example, although some may view the citizenship question as just another demographic variable, in an environment where the Trump White House has heavily primed Hispanic distrust in the government, others may perceive the same question as unfairly targeting specific groups or as discriminatory. Contextual sensitivity is important from a methodological viewpoint: since there are few material benefits to completing most surveys, any increased costs due to contextual sensitivity, even if relatively limited, can lead to non-random increases in item non-response (Berinsky 2004).

The proposed introduction of a citizenship question on the 2020 U.S. Census is an ideal – and particularly important – context within which to study this question. First, any contextual sensitivity could lead to the underreporting of certain groups, thus having dramatic policy repercussions. For example, 132 federal assistance programs depend on accurate Census estimates in order to properly distribute almost \$690 billion (Hotchkiss and Phelan 2017). Second, as some particularly disturbing past examples show (Nobles 2000), the Census often reflects contemporary constructions of race and ethnicity, which means a study of the citizenship question is also inherently a study of the politics surrounding race in the United States.

Although understanding the effect of the citizenship question has important implications for political science research, an upcoming Supreme Court case makes our study especially timely. Prior research (e.g., Brown et al. 2018) has found some evidence the citizenship question may lower census participation, but these findings are derived from observational data and cross-survey comparisons, which are ill-suited for estimating the causal effect of including (or not including) questions that ask about residents’ citizenship. To our knowledge, we are the first to employ a survey experiment that mirrors the form and content of the actual Census which makes our study especially relevant to this important public policy question.

Using a randomized controlled trial (RCT), we find that asking about U.S. citizenship significantly reduces the overall share of questions that respondents answer, with suggestive evidence that the effects are more pronounced among Hispanic respondents. We also find that the citizenship question significantly reduces the number of household members reported as being Hispanic. These patterns are particularly stark among Hispanics

who report being born in Mexico or Central America. Extrapolating our results to the general population, we estimate that asking about citizenship would reduce the share of Hispanics recorded by the Census by approximately 6.07 million, or around 12.03 percent of the 2010 Hispanic population – a sizable reduction in the share of the U.S. population that would be recorded as Hispanic.

Research Design

We designed our survey experiment to match the short form of the U.S. Census, with the questions taken verbatim from the Census’ proposed questions. A third-party vendor (Qualtrics) recruited the survey panel and implemented the study in two waves. The first wave (n = 4,104) began on October 19, 2018 and targeted non-Hispanics (employing an English survey instrument), using self-reported demographic information maintained by Qualtrics. The second wave (n = 4,931) began approximately one week later (on October 25, 2018) and targeted Hispanics (using English and Spanish survey instruments) in order to facilitate meaningful subgroup inferences. In Section S2 of the SI, we report demographic breakdowns for both waves of our survey.

To evaluate the impact of asking about household members’ citizenship on item non-response and response quality, we randomly assigned half of the respondents (n = 4,497) to receive a “Citizenship Treatment” in which we asked, for each member of their household, “Is this person a citizen of the United States?”. The other half (n = 4,538) did not receive the citizenship question for any household member. Because the Census Bureau has not yet indicated where the citizenship question will be located within the 2020 Census, we also randomly rotated the order in which the citizenship question appeared, conditional on the household member in question.

An obvious difference between our study and the actual U.S. census is our status as academic researchers, which might lead to confidence among respondents that data would not be used for immigration purposes. To assess this, we also randomly assigned half of the respondents (n = 4,454) to receive a “Census Prompt” treatment, independently of the first randomization, consisting of a short note at the bottom of their consent form saying “Your responses will be shared with the U.S. Census Bureau,” and requiring respondent consent. The other half (n = 4,581) received no prompt. (Additional details on survey logistics can be found in Sections S1-S2 of our [Supplemental Information](#) or SI.)

Results

Table 1: Treatment Effects on Item Non-Response

	Treatment	Control	Difference	<i>t</i> -statistic	<i>p</i> -value
All	26.22	23.15	3.07	3.956	> 0.001
Hispanic	34.80	30.59	4.21	3.494	> 0.001
<i>Mexico/Central America</i>	20.97	9.93	11.04	3.298	0.001
<i>Puerto Rico/Cuba</i>	13.56	11.78	1.78	0.566	0.572
Non-Hispanic	17.59	15.44	2.15	2.360	0.018

Note: Treatment mean, control mean, and the difference between the two are shown in the first three columns. Last two columns report results from two-sample *t*-tests.

We now turn to our core results concerning treatment effects on item non-response and the underreporting of Hispanic household members. These results are outlined in Tables 1 and 2.

Treatment Effects on Item Non-Response

Beginning with Table 1, we operationalize survey item non-response as the percent of the survey questions for which the respondent submitted a response. Since our Citizenship Treatment was not introduced until Q5, we only consider questions appearing after this question when assessing treatment effects. Using this measure, we find receiving the Citizenship Treatment increases the overall share of questions skipped by 3.07 percentage points (*t*-statistic = 3.956, *p*-value less than 0.001). We find the Census Prompt does not significantly affect the share of questions skipped after Q5 (*t*-statistic = 0.063, *p*-value = 0.950). However, in the SI, we show the Census Prompt does significantly increase the percent of questions skipped in Q1-Q4 (*t*-statistic = 2.322, *p*-value = 0.020). We suspect that after the citizenship question – where we find consistent results throughout our study – appears, its relatively greater salience likely overwhelms the effect of the Census Prompt.

Consistent with our discussion of contextual sensitivity, we also find suggestive evidence that this effect was more pronounced for Hispanics, who skipped 4.21 points more of the questions after the Citizenship Treatment was introduced (*t*-statistic = 3.494, *p*-value is less than 0.001). Given that the current administration has disproportionately targeted its anti-immigrant

policies and rhetoric towards Hispanics from Mexico and Central America, in our pre-analysis plan we predicted more pronounced treatment effects for Hispanics originating from these countries. As anticipated, for these respondents, after receiving the Citizenship Treatment the percent of questions skipped increased by a much larger 11.04 percentage points (t-statistic = 3.298, p-value = 0.001). Since Hispanics who originate from Puerto Rico and Cuba tend to be U.S. Citizens, we also pre-registered this subgroup as an important point of comparison. As also anticipated, the corresponding effect among Hispanics who listed Puerto Rico or Cuba as their birth country was far smaller: 1.78 percentage points (t-statistic = 0.566, p-value = 0.572). We also found a smaller difference of 2.15 percentage points for non-Hispanics (t-statistic = 2.360, p-value = 0.018).

Treatment Effects on Household Members' Race/Ethnicity

A key concern is whether asking about citizenship would affect respondents' willingness to report important demographic information, like household members' race or ethnicity, especially given the current political environment which may make such questions contextually sensitive. Conditioning on the number of household members initially reported by the respondent, the Citizenship Treatment is associated with a 3.25 percentage point increase in the share of questions skipped about members' race/ethnicity (t-statistic = 4.131, p-value less than 0.001). The corresponding effect of the Citizenship Treatment among Hispanics is more pronounced: a 4.86-point increase (t-statistic = 3.951, p-value less than 0.001).

As predicted, we again see even larger, significant effects for Hispanics listing Mexico or a country in Central America as their country of birth. Here, the Citizenship Treatment is associated with a 13.81-point increase in skipped race/ethnicity questions (t-statistic = 3.369, p-value less than 0.001). Finally, among Hispanics who listed either Puerto Rico or Cuba as their country of birth we do not find a significant effect (t-statistic = 0.530, p-value = 0.597). We also see a smaller and insignificant difference of 1.81 percentage points for non-Hispanics (t-statistic = 1.961, p-value = 0.50).

Treatment Effects on Household Members' Ages

Another concern is the possibility that information about citizenship status could be used against families with non-citizen children. We thus

also consider whether the Citizenship Treatment leads to more questions skipped regarding household members' ages. Conditional on the number of household members initially reported, respondents who received the Citizenship Treatment are significantly more likely to skip the questions concerning household members' ages, on average by 3.32 percentage points (t-statistic = 4.111, p-value less than 0.001). We see suggestive evidence that the effect is stronger among Hispanics, who experience a larger, statistically significant increase in questions skipped of 4.56 percentage points (t-statistic = 3.597, p-value less than 0.001).

Once again, as expected, we find even larger, and significant, effects for Hispanic respondents who report being born in either Mexico or a Central American country. For these respondents the Trump administration's rhetoric and policies regarding immigrants and immigration appears, as expected, to have made questions regarding household members' ages contextually sensitive, leading to fewer responses when the citizenship question is introduced. We ultimately find the Citizenship Treatment is associated with a 10.95-point increase in the number of questions related to age and date-of-birth skipped (t-statistic = 3.274, p-value = 0.001). Among Hispanics who listed either Puerto Rico or Cuba as their country of birth we do not find a significant effect (t-statistic = 0.045, p-value = 0.964). Finally, we again see a smaller difference of 2.27 percentage points for non-Hispanics (t-statistic = 2.423, p-value = 0.015).

Treatment Effects on Percent of Household Reported as Being Hispanic

Table 2: Treatment Effects on Percent of Household Reported as Being Hispanic

	Treatment	Control	Difference	t-statistic	p-value
All	31.01	35.04	-4.03	4.224	> 0.001
Hispanic	53.43	59.38	-5.95	4.359	> 0.001
<i>Mexico/Central America</i>	75.35	83.67	-8.32	1.932	0.055
<i>Puerto Rico/Cuba</i>	81.33	85.74	-4.41	1.077	0.283
Non-Hispanic	8.43	9.81	-1.38	1.664	0.096

Note: Treatment mean, control mean, and the difference between the two are shown in the first three columns. Last two columns report results from two-sample t-tests.

Because accurately counting racial/ethnic minorities has substantial implications for federal resource allocations, in Table 2 we consider the effect of the citizenship question on the share of household members

identified by the respondent as being of “Hispanic, Latino, or Spanish Origin.” That is, we consider the percent of household members identified as Hispanic (as opposed to other ethnicities or non-responses) by each respondent. Since question order is randomized, we only consider household members whose race/ethnicity is assigned by the respondent after our Citizenship Treatment is introduced.

Table 2 shows – again consistent with our discussion of contextual sensitivity – those receiving the Citizenship Treatment reported fewer Hispanic household members (31.01 percent of households) compared to those in the control condition (35.04, t- statistic = 4.244, p-value less than 0.001). Hispanic respondents receiving the Citizenship Treatment were 5.95 percentage points fewer household members of Hispanic origin than their counterparts in the control conditions (59.38 vs. 53.43, t-statistic = 4.359, p-value less than 0.001). The corresponding difference among non-Hispanic respondents is a less significant 1.38 points (8.43 vs. 9.81, t-statistic = 1.664, p-value = 0.096).

We again see larger, significant effects for Hispanics listing Mexico or a country in Central America as their country of birth. Here, respondents receiving the Citizenship Treatment reported 8.32 percentage points fewer household members of Hispanic origin (75.35 percent, compared to 83.67 percent in the control condition; t-statistic = 1.932, p-value = 0.055). Once again, among Hispanics who listed either Puerto Rico or Cuba as their birth country, the corresponding difference is smaller (4.41 points; 81.33 vs. 85.74 in the control condition) and insignificant (t-statistic = 1.077, p-value = 0.283).

Robustness Checks

In the SI, we also conduct several robustness checks. First, because the Census Bureau imputes some missing data we replicate our results after separately imputing respondents in the treatment and control conditions. For example, in the imputed control condition, Hispanic respondents report that 73.02 percent of their household members are Hispanic. When receiving the Citizenship Treatment, these same respondents report 64.45 percent of their household members are Hispanic, a difference of 8.57 percentage points (t-statistic = 6.891, p-value less than 0.001). Second, we also replicate our main results using respondents who did not receive the Census Prompt. All of our results hold using this subgroup. For instance, for Hispanic respondents who did not receive the Census Prompt, receiv-

ing the Citizenship Treatment significantly decreases the percentage of household members reported as being Hispanic by 5.67 percentage points (t-statistic = 2.978, p-value = 0.003).

Extrapolating to the 2020 U.S. Census Context

We now consider what, more precisely, our study might predict for the 2020 U.S. Census context in terms of the share of Hispanics that might be undercounted. Our survey purposefully oversampled Hispanics (51.10 percent of our sample) relative to the U.S. population (16.35 percent, as reported by the 2010 U.S. Census). To produce more nationally representative estimates, we created post-stratification weights using a raking algorithm, based on all available information about our respondents (their race/ethnicity, whether they were Democratic identifiers, and their state of residence, based on zip codes). (We provide more details in Section S2-S3 of the SI.)

Applying the estimated national-level treatment effect to the U.S. population, as reported by the 2010 U.S. Census (308,745,538), we estimate that asking about citizenship will reduce the number of Hispanics reported in the 2020 Census by 6,072,068 or 12.03 percent of the 2010 Hispanic population (50,477,594). The 95-percent confidence interval surrounding our estimate is 5,761,284 to 6,382,820, which represents a decrease of 11.41 to 12.64 percent relative to the 2010 Hispanic population. In 2016, before the current administration came to power and initiated its sustained pattern of severe anti-immigrant rhetoric and policies, six percent of all respondents and 7.40 percent of Hispanic respondents did not respond to the ACS citizenship question (Brown et al. 2018), which suggests that our estimates are reasonable, though admittedly suggestive.

Although we cannot say with certainty how the predicted undercounted Hispanics will be distributed, in some configurations congressional apportionment could be affected, with many worrying Texas and California may lose seats (Lind 2018). In Table S6 of the SI we estimate the effect of the citizenship question across all 435 districts. There we find in some districts where Hispanics represent a larger share of the population – especially in Southern California – our Citizenship Treatment would have a larger impact, which suggests apportionment could also be affected by introducing the citizenship question.

Discussion and Conclusion

This study presents the first explicit causal exploration of the impact of asking about citizenship on the 2020 U.S. Census. We find that asking about citizenship status significantly increases the percent of questions skipped, with particularly strong effects among Hispanics, and makes respondents less likely to report having Hispanic household members. Aggregating this to the national level suggests that asking a citizenship question may lead to an undercounting of Hispanics of between 5,761,284 and 6,382,820 in the 2020 Census (based on 2010 figures).

We note two caveats regarding our study. The first is that we likely underestimate the effect of asking about citizenship status on the 2020 Census. Not only are we university affiliated academic researchers – and not the U.S. Government – and so respondent concerns over providing the government with personal information may not have inhibited participation in this survey, but our respondents were paid panelists and thus financially incentivized to complete the survey. The second is that our study was ill suited for estimating the causal effects of citizenship questions on household member undercounts. In order to mirror the actual census form, all citizenship questions appeared in our survey after respondents listed the number of members of their household. Additionally, unlike the Census Bureau, we had no “baseline truth” against which to compare reported household size. Future research might be better equipped to assess the potential impact of asking about citizenship status on household size reporting.

These caveats aside, the key takeaway is that including a citizenship question will likely result in undercounts of Hispanics and this is at least partially attributable to simple item non-response. Whether the source is breaking off from the survey or withholding information about some household members, item non-response is one of the main concerns surrounding the citizenship question (Lind 2018). A single respondent answers questions about everyone in the household, which means that any false or incomplete information about certain household members is equivalent to those individuals not responding to the survey. This is why a major part of the U.S. Census Bureau’s non-response follow-up operation focuses on contacting households that returned incomplete surveys (Walker et al. 2010).

Scholars have paid relatively little attention to the risks of item non-response introduced by including the citizenship question on the 2020 Census. Our study thus makes an important addition to the literature. In doing so, we provide a clear – and politically important – cautionary example for

survey researchers. Whether researchers are administering the U.S. Census, American National Election Study (ANES), or simple experiments at their home institutions, questions about race and ethnicity are inherently political, and thus contextually sensitive, especially in the current political environment. Researchers should therefore take considerable care in formulating those and other demographic questions. Otherwise, survey response rates could be negatively affected. In this way, we provide a key foundation for future work on the importance of contextual sensitivity to survey research and studies of American politics, more broadly.

Acknowledgements

This research was supported by the Shorenstein Center on Media, Politics, and Policy at the John F. Kennedy School of Government, Harvard University. We thank Rafael Carbonell, Kyla Fullenwider, Candelaria Garay, Sunshine Hillygus, Horacio Larreguy, David Lazer, Michelle Lyons, O'Reilly Miami, Nicco Mele, Noelia Negri, Nathan Richard, and John Thompson for helpful comments and suggestions. We also thank Amaris Taylor for her help at Qualtrics. Liz Schwartz, Rebecca Roach, Carmen Nobel, Denise Ordway, and John Sides also helped immensely with the public facing versions of our paper.

References

Supplemental Information available here: https://shorensteincenter.org/wp-content/uploads/2019/03/census_supplemental_information.pdf

Berinsky, Adam J. 2004. *Silent Voices: Opinion Polls and Political Representation in America*. Princeton, NJ: Princeton University Press.

Brown, J. David et al. 2018. *Understanding the Quality of Alternative Citizenship Data Sources for the 2020 Census*. Technical Report. U.S. Census Bureau.

Bureau, U.S. Census. 2017. *Race & Ethnicity*. Technical Report. U.S. Census Bureau.

De Leeuw, Edith D, Joop J Hox and Mark Huisman. 2003. "Prevention and Treatment of Item Nonresponse." *Journal of Official Statistics* 19(2):153–176.

Hotchkiss, Marisa and Jessica Phelan. 2017. *Uses of Census Bureau Data in Federal Funds Distribution: A New Design for the 21st Century*. Technical Report. U.S. Census Bureau.

Lind, Dara. 2018. "The Citizenship Question on the 2020 Census, Explained." *Vox*.

Nobles, Melissa. 2000. "History Counts: A Comparative Analysis of Racial/Color Categorization in US and Brazilian Censuses." *American Journal of Public Health* 90(11):1738-1745.

Walker, Shelley et al. 2010. "2010 Census Nonresponse Followup Operations Assessment." Technical Report. U.S. Census Bureau.

Supporting Information for:
Estimating the Effect of Asking About Citizenship on
the U.S. Census: Results from a Randomized Controlled
Trial

Contents

S1 Materials and Methods	S2
S1.1 EGAP and IRB	S2
S1.2 Sampling Design	S2
S1.3 Experimental Design	S2
S1.4 English Survey	S4
S1.5 Spanish Survey	S14
S2 Demographic and Balance Statistics	S23
S2.1 Demographics	S23
S2.2 Balance Statistics	S27
S3 Additional Analyses	S30
S3.1 Survey Weights	S30
S3.2 Census Opinion Questions	S31
S3.3 Additional Hypotheses	S33
S3.4 Congressional District Analysis	S36
S3.5 Imputation Analysis	S37
S3.6 Direct Costs Analysis	S38
S3.7 Modeling Proportions	S39

S1 Materials and Methods

S1.1 EGAP and IRB

Our experimental design was pre-registered with EGAP (Evidence in Governance and Politics) and approved on October 23, 2018 (ID# 20181016AA). The Harvard Institutional Review Board (IRB) questioned whether our Census prompt treatment was deception. Since the nature of the study meant that we were open to sharing data with pertinent public agencies and other relevant actors both inside and outside government, our treatment does not constitute deception. We have also arranged to share our findings with several former Census Bureau staff members, who, in turn, have indicated that they will share the data with current Census staff. In addition, all data (de-identified) would be made publicly available, per standard practices within social science. With these assurances, the Harvard IRB agreed to approve our proposed design and assigned our study the following protocol numbers: IRB18-1445, MOD18-1445-01, and MOD18-1445-02.

S1.2 Sampling Design

A third-party vendor (Qualtrics) recruited the survey panel, which included actively managed research panels and panelists drawn from third-party vendors. We chose this approach partly because, compared to using an established polling firm, it reduced the possibility that we were surveying repeated or professional survey takers, which would be a bad match for replicating a census that surveys all U.S. residents. The panel was purposefully not restricted to U.S. citizens (just residents aged 18 or over).

The survey was conducted in two waves, each with an identical incentive structure. The first wave targeted non-Hispanics (using an English-only survey instrument), using self-reported demographic information maintained by Qualtrics. This group ($n = 4,104$) included 3,413 whites, 246 African Americans, 55 Hispanics, 181 Asians/Asian-Americans, 92 members of other racial groups, and 117 respondents who did not identify their race to Qualtrics. (Additional details regarding the demographic composition of our survey respondents can be found in Section S2 of the SI.)

The second wave targeted Hispanics with the goal of obtaining large enough numbers to make meaningful subgroup inferences. Moreover, because some Hispanics are primarily Spanish-speakers, second-wave respondents had the option of taking the survey either in English or Spanish. (Both versions are included in Section S1 of the SI.) The second wave included 4,931 respondents, which included 4,562 Hispanics, 13 whites, 3 African Americans, 1 Asian/Asian-American, 3 members of other racial groups, and 349 respondents who did not identify their race.

S1.3 Experimental Design

The survey was administered online via the Qualtrics platform and included at least 16 questions (or 17 for Hispanics). Our internet-only format follows the anticipated protocol of the 2020 U.S. Census, the bulk of which is expected to take place online. Following the Census format, the survey

first asks about the number of individuals in the respondent’s household as of a certain date, those individuals’ names (on our survey, their initials), the tenure status of the household, and which individual owns the home or pays rent. The survey then asks six demographic questions for each additional household member, in the order listed.

Household Member Citizenship Question (“Citizenship Treatment”). To evaluate the impact of asking about household members’ citizenship on item non-response and response quality, we randomly assigned half of the respondents ($n = 4,497$) to a treatment condition in which we asked, for each member of their household, “Is this person a citizen of the United States?” Similar to the proposed 2020 Census question, we provided five answer categories, which are outlined in Section S1 of the SI. The other half ($n = 4,538$) were randomly assigned to a control condition in which they did not receive the citizenship question for any household member.

Within the subset of questions about household members, the exact placement of the citizenship question on the actual census is yet to be determined. Thus, for our survey, we randomly rotated the order in which the citizenship question appeared, conditional on the household member being asked about.

Data Sharing with Census Treatment (“Census Prompt” Treatment). Any treatment effect associated with a citizenship question may be smaller than if asked by the U.S. Census itself, since individuals might fear reprisal from the federal government but not from academic researchers with no government affiliations. We would thus expect smaller treatment effects associated with the citizenship treatment in our survey as opposed to the actual U.S. Census.

We assess this with another randomized intervention. Specifically, we randomly assigned half of respondents ($n = 4,454$) to receive, independently of the first randomization, a short note at the bottom of their consent form saying “Your responses will be shared with the U.S. Census Bureau.” In order to proceed, we required that respondents indicate agreement. The other half ($n = 4,581$) received no prompt. This priming more closely mirrors the real-world scenario of the U.S. Census, in which respondents might expect that their data are held by a federal government agency (the Census Bureau), with any accompanying personal risks. (Additional details regarding our experimental design can be found in Section S1 of the SI.)

Table S1: Respondents Randomly Assigned to Each Treatment

	Did Not Receive Citizenship Question	Received Citizenship Question	Total
Did Not Receive Census Prompt	2335	2246	4581
Received Census Prompt	2203	2251	4454
Total	4538	4497	9035

Note: This table shows our 2×2 experimental design. The number of respondents randomly assigned to each of our treatments is included in each cell.

As Table S1 summarizes, the final experiment was a 2×2 design, with respondents having an equal probability of assignment to any of the four conditions:

1. No questions on citizenship, no prompt about information sharing (n = 2,335)
2. Citizenship Treatment only (n = 2,246)
3. Census Prompt treatment only (n = 2,203)
4. Citizenship Treatment and Census Prompt treatment (n = 2,251)

Of 9,035 respondents, 4,497 received the citizenship question, statistically indistinguishable from the expected 0.50 proportion ($\chi^2 = 0.177$, $df = 1$, p -value = 0.674). Of 9,035 respondents, 4,454 received the census prompt, a figure that is also statistically indistinguishable from the expected proportion ($\chi^2 = 1.757$, $df = 1$, p -value = 0.185). In the next subsections we provide all the questions we used in both the English and Spanish versions of our survey.

S1.4 English Survey

Our survey was constructed to mirror the Internet Self-Response (ISR) for the 2020 Census. Similar to the paper version, the first questions are about the respondent's household: the number of individuals living there as of a certain date (on the 2020 Census, it will be April 1, 2020, and on our survey, we used September 1, 2019), those individuals' names (on our survey, their initials), the tenure status of the household, and which individual owns the home or pays the rent (Person 1). Then the demographic questions are asked about each of the individuals in the order that the survey-taker listed them.

In addition to the basic structure of the survey, we also attempted to mirror the same look, feel, and skip logic of the ISR. For example, none of the demographic questions were forced-responses which means respondents could skip those questions and still receive the survey incentive. Below we provide the consent form we used for all survey respondents. At the bottom of this form is the Census prompt treatment. The questions we asked in the English version appear shortly thereafter. In those questions, the citizenship question appears as Question 5.

Consent Form

Key Information

The following is a short summary of this study to help you decide whether or not to be a part of this study. More detailed information is listed later on in this form.

Why am I being invited to take part in a research study?

We invite you to take part in a research study because you are a U.S. resident over 18. What should I know about a research study? Someone will explain this research study to you. Whether or not you take part is up to you. Your participation is completely voluntary. You can choose not to take part. You can agree to take part and later change your mind. Your decision will not be held against you. You can ask all the questions you want before you decide.

Why is this research being done?

This research is being done to estimate the response rates to surveys that include different types of questions.

How long will the research last and what will I need to do?

We expect that you will be in this research study for approximately seven to twenty minutes. You will be asked to complete a survey about your household and about some opinions on social issues.

Is there any way being in this study could be bad for me?

We don't believe there are any risks from participating in this research.

Will being in this study help me in any way?

There are no benefits to you from your taking part in this research. We cannot promise any benefits to others from your taking part in this research. However, possible benefits to others include more informed policymaking.

Detailed Information

The following is more detailed information about this study in addition to the information listed above.

What is the purpose of this research?

The purpose of this research is to learn more about what types of survey questions individuals are and are not comfortable responding to. Many institutions which administer surveys are concerned about the rate of individuals dropping out of surveys, because they are not able to learn as much information when individuals drop out of their surveys. Therefore, it is important to understand as much as possible about what types of survey questions lead to individuals dropping out of surveys.

How long will I take part in this research?

You will be asked to take an online survey one time, and the survey will take between five and fifteen minutes.

What can I expect if I take part in this research?

You can expect to take an online survey from your computer or mobile device that will take between seven and twenty minutes. This survey will ask you some questions about yourself and some questions about your attitudes.

What happens if I say yes, but I change my mind later?

You can leave the research at any time it will not be held against you.

If I take part in this research, how will my privacy be protected? What happens to the information you collect?

This study will not collect any personally identifying information. Nevertheless, efforts will be made to limit the use and disclosure of your Personal Information, including research study and medical records, to people who have a need to review this information. We cannot promise complete secrecy. Organizations that may inspect and copy your information include the IRB and other representatives of this organization, as well as the Harvard Kennedy School of Government and other representatives of this organization.

What else do I need to know?

You will be compensated the amount you agreed upon before you entered into the survey.

Who can I talk to?

If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team at:

[NAME 1]: [EMAIL] or [PHONE NUMBER]

[NAME 2]: [EMAIL] or [PHONE NUMBER]

[NAME 3]: [EMAIL] or [PHONE NUMBER]

This research has been reviewed and approved by the Harvard University Area Institutional Review Board (“IRB”). You may talk to them at (617) 496-2847 or cuhs@harvard.edu if: Your questions, concerns, or complaints are not being answered by the research team. You cannot reach the research team. You want to talk to someone besides the research team. You have questions about your rights as a research subject. You want to get information or provide input about this research.

You may download a copy of this information for your records by clicking here.

[*This is the Census prompt treatment. Randomly assigned to approximately half of the respondents. All others do not receive this question.*] Your responses will be shared with the U.S. Census Bureau:

I understand that my responses will be shared with the U.S. Census Bureau

Questions

1. How many people were living or staying in this house, apartment, or mobile home on September 1, 2018?
- 2(a). Were there any additional people staying here on September 1, 2018 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
- 2(b). [*If R answered 2(a) with any answer except “No additional people”.*] How many additional people?
3. Is this house, apartment, or mobile home –
 - Owned by you or someone in this household with a mortgage or loan?
 - Owned by you or someone in this household free and clear (without a mortgage or loan)?
 - Rented?
 - Occupied without payment of rent?

[*Beginning here, questions are asked for the number of household members listed in Question 1. The section always starts with Question 4(a), then Questions 5*-10 are randomized. After the last randomized question is asked for Person 1, then Question 4(a) is asked for the next household member. For Person 2 and above we add Questions 4(b)* and 4(c)* to the randomized questions (5*-10) and the section repeats.*]

Household Demographics Instructions: 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.

4(a). What are Person X's initials? *Print initials below*

First initial:

MI:

Last initial:

5*. [This is the citizenship question treatment. Randomly assigned to approximately half of the respondents. All others do not receive this question.] Is this person a citizen of the United States?

- 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

6. What is this person's age? *For babies less than 1 year old, do not write the age in months. Write 0 as the age.*

7. What is this person's year of birth?

8. Is this person of Hispanic, Latino, or Spanish origin?

- No, not of Hispanic, Latino, or Spanish origin
- Yes, Mexican, Mexican Am., Chicano
- Yes, Puerto Rican
- Yes, Cuban
- Yes, another Hispanic, Latino, or Spanish origin – *Print, for example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc.*

9. What is this person's race or ethnicity? *Select all boxes that apply and/or enter details as necessary. Note, you may report more than one group.*

- | | | |
|---|--|---------------------------------------|
| • German | • African American | <i>South African, Barbadian, etc.</i> |
| • Irish | • Jamaican | |
| • English | • Haitian | • Chinese |
| • Italian | • Nigerian | • Filipino |
| • Polish | • Ethiopian | • Asian Indian |
| • French | • Somali | • Vietnamese |
| • Other White – <i>Print, for example, Scottish, Norwegian, Dutch, etc.</i> | • Other Black – <i>Print, for example, Ghanaian,</i> | • Korean |
| | | • Japanese |

- Other Asian – *Print, for example, Pakistani, Cambodian, Hmong, etc.*
- American Indian – *Print, for example, Navajo Nation, Blackfeet Tribe, Muscogee (Creek) Nation, etc.*
- Alaskan Native – *Print, for example, Native Village of Barrow Inupiat Traditional Government, Tlingit, Orustaramuit Native Village, etc.*
- Central or South American Indian – *Print, for example, Mayan, Aztec, Taino, etc.*
- Lebanese
- Iranian
- Egyptian
- Syrian
- Moroccan
- Israeli
- Other Middle Eastern or North African – *Print, for example, Algerian, Iraqi, Kurdish, etc.*
- Native Hawaiian
- Samoan
- Chamorro
- Tongan
- Fijian
- Marshallese
- Other Pacific Islander – *Print, for example, Palauan, Tahitian, Chuukese, etc.*
- Some other race – *Print race or origin.*

10. What is this person's sex?

- Male
- Female

4(b).* [Beginning with Person 2, we then add the following questions. These always appear after Question 4(a) and are randomized with Questions 5, 6, 7, 8, 9, and 10.]

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

4(c).* How is this person related to Person 1?

- 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

[Beginning here, repeat sex, age, Hispanic origin, race, and citizenship questions for all enumerated household members.]

11. Have you ever heard of the United States Census, or have you not heard of this?

- I have heard of the United States Census
- I have not heard of the United States Census

12(a). How likely are you to participate in the 2020 United States Census? By participate, we mean fill out and mail in a Census form or fill one out online. Would you say you...

- Definitely will
- Probably will
- Might or might not
- Probably will not
- Definitely will not

12(b). *[If R answered 12(a) with “Might or might not,” “Probably will not,” or “Definitely will not”.]* By participate, we mean fill out and mail in a Census form or fill one out online. Would you say someone else in your household...

- Definitely will
- Probably will
- Might or might not
- Probably will not

- Definitely will not
13. How important do you think the Census is for the United States? Would you say it is...
- Very important
 - Somewhat important
 - Not too important
 - Not at all important
 - Don't know enough to say
14. Do you believe that answering and sending back or completing online your United States Census form would...
- Personally benefit you
 - Personally harm you
 - Neither benefit or harm you
 - Don't know enough to say
- 15(a). Do you believe that answering and sending back or completing online your United States Census form would...
- Benefit your community
 - Harm your community
 - Neither benefit or harm your community
 - Don't know enough to say
- 15(b). *[If R answered 15(a) with "Benefit your community" or "Harm your community". Answer is piped into this question.]* Why do you say the Census would [benefit/harm] your community?
16. How concerned are you, if at all, that the Census Bureau not keep answers to the 2020 Census confidential?
- Extremely concerned
 - Very concerned
 - Somewhat concerned
 - Not too concerned
 - Not at all concerned
 - Don't know enough to say
17. How concerned are you, if at all, that the Census Bureau will share answers to the 2020 Census with other government agencies?

- Extremely concerned
- Very concerned
- Somewhat concerned
- Not too concerned
- Not at all concerned
- Don't know enough to say

18. How concerned are you, if at all, that the answers you provide to the 2020 Census will be used against you?

- Extremely concerned
- Very concerned
- Somewhat concerned
- Not too concerned
- Not at all concerned
- Don't know enough to say

19(a). Do you think the results of the United States Census help one political party (the Republican Party or the Democratic Party) more than the other, or don't you think so?

- Yes
- No
- Don't know enough to say

19(b). *[If R answered 19(a) with "Yes".]* Which political party do you think the United States Census helps more?

20. As far as you know, is the Census used to determine whether someone is in this country legally, or is it not used for this?

- Yes, it is used to determine whether someone is in this country legally
- No, it is not used to determine whether someone is in this country legally
- Don't know enough to say

21. As far as you know, is the Census used to decide how many representatives each state will have in Congress, or is it not used for this?

- Yes, it is used to decide how many representatives each state will have in Congress
- No, it is not used to decide how many representatives each state will have in Congress
- Don't know enough to say

22. As far as you know, is the Census used to decide how much money communities will get from the government, or is it not used for this?
- Yes, it is used to decide how much money communities will get from the government
 - No, it is not used to decide how much money communities will get from the government
 - Don't know enough to say
23. As far as you know, is the Census Bureau supposed to keep the personal information you provide on the 2020 Census form confidential, or are they not supposed to do that?
- Yes, it is supposed to keep the personal information you provide on the 2020 Census form confidential
 - No, it is not supposed to keep the personal information you provide on the 2020 Census form confidential
 - Don't know enough to say
24. *[If R identified any household member as being of Hispanic, Latino, or Spanish Origin (see Question 8).]* Have you seen or heard anything recently from Hispanic/Latino civic, religious, media or community groups encouraging or discouraging you to from filling out your 2020 Census form?
- Yes
 - No

S1.5 Spanish Survey

For the Spanish translation, we used the U.S. Census Bureau's own translation as much as possible. For any questions that did not appear on the Census short form, including our university-required consent form and debriefing materials, we obtained translations from a professional translation company, which we then vetted with the assistance of several native Spanish speakers. Only 210 respondents took the survey in Spanish.

Consent From

Información clave

Esta es una breve explicación del estudio para ayudarlo a decidir si usted desearía o no participar en el mismo. Incluimos información más detallada a continuación en este formulario.

¿Por qué fui invitado a participar en este estudio?

Lo invitamos a participar por el hecho de ser residente de los Estados Unidos y ser mayor de 18 años.

¿Qué debo saber sobre este estudio?

- Alguien le va a explicar este estudio.
- Participar o no en este estudio es su decisión.
- Su participación es completamente voluntaria.
- Puede elegir no participar.
- Puede participar y luego cambiar de opinión.
- Su decisión no se utilizará en su contra.
- Puede hacer todas las preguntas que desea antes de decidir si quiere participar.

¿Por qué se está haciendo esta investigación?

Esta investigación se está haciendo para medir las tasas de respuesta a encuestas que incluyen diferentes tipos de preguntas.

¿Cuánto durará la investigación y qué debo hacer?

Estimamos que participar en este estudio le tomará entre siete y veinte minutos. Se le pedirá completar una encuesta sobre su hogar y sus opiniones sobre asuntos sociales.

¿Puede este estudio perjudicarme de alguna manera?

No creemos que haya ningún riesgo de participar en este estudio.

¿Participar en este estudio me ayudará de alguna manera?

No hay beneficios para usted por participar en este estudio. Tampoco podemos prometer beneficios a otras personas como resultado de su participación en el mismo. Sin embargo, un posible beneficio para los demás puede ser que las decisiones políticas se tomen de una manera más informada.

Información detallada

A continuación se incluye información más detallada sobre este estudio además de lo previamente mencionado.

¿Cuál es el propósito de esta investigación?

El propósito de este estudio es aprender más sobre los tipos de preguntas en encuestas que las personas se sienten cómodas o incómodas en responder. A muchas instituciones que hacen encuestas les preocupa que las personas abandonen las mismas, ya que no pueden obtener suficiente información cuando los individuos abandonan sus encuestas. Por ello, es importante saber los más que se pueda sobre los tipos de preguntas hacen que las personas abandonen las encuestas.

¿Cuánto tiempo me tomará participar?

Se le solicitará que responda una encuesta en línea que le llevará entre siete y veinte minutos.

¿Qué puedo esperar si participo en este estudio?

Usted completará una encuesta desde su computadora o teléfono móvil. Esto le tomará entre siete y veinte minutos. La encuesta incluirá preguntas acerca de usted y algunas preguntas sobre sus opiniones.

¿Qué pasa si digo que sí, pero luego cambio de opinión?

Puede abandonar la investigación en cualquier momento en que no se llevará a cabo en su contra.

¿Si participo en el estudio, ¿cómo se protegerá mi privacidad? ¿Qué pasa con la información recolectada?

Este estudio no recopilará ningún tipo de información que permita identificarle personalmente. No obstante, se realizarán esfuerzos para limitar el uso y divulgación de su información personal, incluyendo los datos de la investigación y registros médicos, únicamente a las personas que necesiten acceder a la misma. No podemos prometer la ocultación completa de la información. Entre las organizaciones que podrían inspeccionar una copia de su información se encuentra la Junta de Revisión Institucional (IRB) de la Universidad de Harvard y otros representantes de la misma; así como la Escuela de Gobierno John F. Kennedy de la Universidad de Harvard y otros representantes de ésta.

¿Qué más necesito saber?

Será compensado por el monto acordado antes de que comience a responder la encuesta.

¿Con quién puedo hablar?

Si tiene preguntas, inquietudes o quejas, o cree que el estudio lo ha afectado de alguna manera comuníquese con el equipo de investigación:

[NAME 1]: [EMAIL] ó [PHONE NUMBER]

[NAME 2]: [EMAIL] ó [PHONE NUMBER]

[NAME 3]: [EMAIL] ó [PHONE NUMBER]

Este estudio ha sido revisado y aprobado por la Junta de Revisión Institucional de la Universidad de Harvard. Puede comunicarse con ellos al número telefónico (617) 496-2847 o por correo electrónico a cuhs@harvard.edu en los siguientes casos:

- El equipo de investigación no responde a sus dudas, inquietudes o quejas.
- No puede comunicarse con el equipo de investigación.
- Desea hablar con alguien fuera del equipo de investigación.
- Tiene dudas sobre sus derechos como sujeto de la investigación.
- Desea obtener información o dar su opinión sobre la investigación.

[This is the Census prompt treatment. Randomly assigned to approximately half of the respondents. All others do not receive this question.] Sus respuestas serán compartidas con el U.S. Census Bureau:

Entiendo que mis respuestas serán compartidas con el U.S. Census Bureau

Questions

1. ¿Cuántas personas estaban viviendo o quedándose en esta casa, departamento o casa móvil el primero de septiembre de 2018?
- 2(a). ¿Había personas adicionales quedándose aquí el 1 de septiembre de 2018 que no incluyó en la Pregunta 1? *Indiquen todas que correspondan.*
 - Niños, emparentados o no, tales como bebés recién nacidos, nietos o niños acogidos (foster children)
 - Parientes adultos, tales como hijos mayores de edad, primos o parientes políticos
 - Personas adultas que no sean parientes, tales como compañeros de casa o cuarto, o niñeras que viven en el hogar
 - Personas que se quedan aquí temporalmente
 - No hay personas adicionales
- 2(b). *[If R answered 2(a) with any answer except “No additional people”.]* ¿Cuántas personas adicionales?
3. ¿Es esta casa, departamento o casa móvil?
 - Propiedad suya o de alguien viviendo en esta casa con una hipoteca o crédito hipotecario? *Esto incluye los préstamos con la propiedad como garantía.*
 - Propiedad suya o de alguien en esta casa totalmente pagada y sin deuda (sin una hipoteca o crédito hipotecario)?
 - Alquilado(a) o rentado(a)?
 - Ocupado(a) sin pago de alquiler o renta?

[Beginning here, questions are asked for the number of household members listed in Question 1. The section always starts with Question 4(a), then Questions 5-10 are randomized. After the last randomized question is asked for Person 1, then Question 4(a) is asked for the next household member. For Person 2 and above we add Questions 4(b)* and 4(c)* to the randomized questions (5*-10) and the section repeats.]*

Household Demographics Instructions: Por favor, provea información para cada persona que vive aquí. Si hay alguien que vive aquí que paga el alquiler (renta) o es propietario de esta vivienda, comience la lista con él o ella como la Persona 1. Si el propietario o la persona que paga el alquiler (renta) no vive aquí, comience la lista con cualquier adulto que viva aquí como la Persona 1.

- 4(a). ¿Cuáles son las iniciales de la Persona 1? *Escriba las iniciales a continuación.*

Iniciales de nombre:

Iniciales del apellido(s):

5*. [This is the citizenship question treatment. Randomly assigned to approximately half of the respondents. All others do not receive this question.] ¿Ciudadanía: Es esta persona ciudadana de los Estados Unidos?

- Sí, nació en los Estados Unidos
- Sí, nació en Puerto Rico, Guam, las Islas Vírgenes de los Estados Unidos o las Islas Marianas del Norte
- Sí, nació en el extranjero de padre o madre que es ciudadano(a) de los EE. UU.
- Sí, es ciudadana de los Estados Unidos por naturalización. *Escriba el año de naturalización*
- No, no es ciudadana de los Estados Unidos

6. ¿Edad: Cuál es la edad de esta persona? *Para bebés menores de un año, no escriba los meses de edad. Solo escriba 0.*

7. ¿Cuál es su fecha de nacimiento?

8. ¿Origen hispano: Es esta Persona de origen hispano, latino o español?

- No, no es de origen hispano, latino o español
- Sí, es mexicano, mexicano-americano, chicano
- Sí, es puertorriqueño
- Sí, es cubano
- Sí, es de otro origen hispano, latino o español *Escriba, por ejemplo, salvadoreño, dominicano, colombiano, guatemalteco, español, ecuatoriano, etc.*

9. ¿Raza: Cuál es la raza de esta persona?

- | | | |
|--|---|--|
| <ul style="list-style-type: none"> • Orígenes étnicos BLANCOS • Alemán • Irlandés • Inglés • Italiano • Polaco • Francés • Otro: escriba, por ejemplo, escocés, noruego, holandés, etc. • Orígenes étnicos NEGROS o AFROAMER- | <ul style="list-style-type: none"> ICANOS • Afroamericano • Jaiquirino • Haitiano • Nigeriano • Etíope • Somalí • Otro: escriba, por ejemplo, ghanés, sudafricano, barbadense, etc. • Orígenes étnicos ASIÁTICOS | <ul style="list-style-type: none"> • Chino • Chino • Indio asiático • Vietnamita • Coreano • Japonés • Otro: escriba, por ejemplo, pakistaní, camboyano, hmong, etc. • INDÍGENA DE LAS AMÉRICAS o NATIVO DE ALASKA |
|--|---|--|

- | | | |
|--|---|---|
| <ul style="list-style-type: none"> ● Indígena americano: escriba, por ejemplo, Navajo Nation, Black-foot Tribe, Muscogee (Creek) Nation, etc. ● Nativo de Alaska: Escriba, por ejemplo, Native Village of Barrow Inupiat Traditional Government, Tlingit, Orustaramuit Native Village, etc. ● Indígena de América Central o Sudamérica: escriba, por ejemplo, maya, azteca, taíno, etc. | <ul style="list-style-type: none"> ● Orígenes étnicos del MEDIO ORIENTE O ÁFRICA DEL NORTE ● Libanés ● Iraní ● Egipcio ● Sirio ● Marroquí ● Israelí ● Otro: escriba, por ejemplo, argelino, iraquí, kurdo, etc. ● Orígenes étnicos NATIVO DE HAWÁI | <p style="text-align: center;">o ISLEÑO DEL PACÍFICO</p> <ul style="list-style-type: none"> ● Nativo de Hawái ● Samoano ● Chamorro ● Tongano ● Fiyiano ● de las Islas Marshall ● Otro: escriba, por ejemplo palauano, tahitiano, chuukés, etc. ● Alguna otra raza u origen étnico |
|--|---|---|

10. ¿Sexo: Cuál es el sexo de esta persona?

- Masculino
- Feminino

4(b).* *[Beginning with Person 2, we then add the following questions. These always appear after Question 4(a) and are randomized with Questions 5, 6, 7, 8, 9, and 10.]*

¿Esta persona generalmente vive o se queda en algún otro lugar?

- No
- Sí, para ir a la universidad
- Sí, por una orden militar
- Sí, por un empleo o negocio
- Sí, en un hogar de ancianos o nursing home
- Sí, con el padre, la madre u otro pariente
- Sí, en una vivienda temporal o segunda residencia
- Sí, en una cárcel o prisión
- Sí, por alguna otra razón

4(c).* ¿Relación: Cómo está esta persona relacionada con la Persona 1?

- Esposo(a) del sexo opuesto
- Pareja no casada del sexo opuesto
- Esposo(a) del mismo sexo

- Pareja no casada del mismo sexo
- Hijo(a) biológico(a) de sangre
- Hijo(a) adoptivo(a)
- Hijastro(a)
- Hermano(a)
- Padre o madre
- Nieto(a)
- Suegro(a)
- Yerno o nuera
- Otro pariente
- Roommate o compañero(a) de casa
- Niño(a) acogidos (foster child)
- Otra persona que no es pariente

[Beginning here, repeat sex, age, Hispanic origin, race, and citizenship questions for all enumerated household members.]

11. ¿Ha oído mencionar el Censo de los Estados Unidos, o no lo ha oído mencionar?

- He oído mencionar el Censo de los Estados Unidos
- No he oído mencionar el Censo de los Estados Unidos

12(a). ¿Qué tan probable es que participe en el Censo de los Estados Unidos del año 2020? Por participar nos referimos a que llene y envíe por correo el formulario del Censo o a que llene el formulario en línea. ¿Diría usted que?

- Definitivamente participaré
- Probablemente participaré
- Tal vez participaré
- Probablemente no participaré
- Definitivamente no participaré

12(b). *[If R answered 12(a) with “Might or might not,” “Probably will not,” or “Definitely will not”.]* Por participar nos referimos a llenar y enviar por correo el formulario del Censo o llenar el formulario en línea. ¿Qué tan probable es que alguien más en su vivienda participe en el Censo del año 2020?

- Definitivamente participará
- Probablemente participará

- Tal vez participará
 - Probablemente no participará
 - Definitivamente no participará
13. ¿Qué tan importante cree que el Censo es para los Estados Unidos? ¿Diría que es?
- Muy importante
 - Un poco importante
 - No tan importante
 - No importante
 - No tiene suficiente conocimiento
14. ¿Cree que contestar y enviar su formulario del Censo de los Estados Unidos o llenar el formulario en línea puede?
- Beneficiarle personalmente
 - Perjudicarle personalmente
 - Ni beneficiarle ni perjudicarle a usted personalmente
 - No tiene suficiente conocimiento
- 15(a). ¿Cree que contestar y enviar su formulario del Censo de los Estados Unidos o llenar el formulario en línea puede?
- Beneficiar a su comunidad
 - Perjudicar a su comunidad
 - Ni beneficiar ni perjudicar a su comunidad
 - No tiene suficiente conocimiento
- 15(b). *[If R answered 15(a) with “Benefit your community” or “Harm your community”. Answer is piped into this question.]* ¿Por qué dijo que el Censo podría beneficiar/ perjudicar a su comunidad?
16. ¿Qué tan preocupado(a) está que el Census Bureau (La Oficina del Censo) no mantendrá confidenciales sus respuestas al Censo del 2020?
- Profundamente preocupado(a)
 - Muy preocupado(a)
 - Algo preocupado(a)
 - No tan preocupado(a)
 - Nada preocupado(a)
 - No tiene suficiente conocimiento

17. ¿Qué tan preocupado(a) está que el Census Bureau (La Oficina del Censo) compartirá sus respuestas al Censo del 2020 con otras agencias gubernamentales?
- Profundamente preocupado(a)
 - Muy preocupado(a)
 - Algo preocupado(a)
 - No tan preocupado(a)
 - Nada preocupado(a)
 - No tiene suficiente conocimiento
18. ¿Qué tan preocupado(a) está que sus respuestas al Censo del 2020 se usarán en su contra?
- Profundamente preocupado(a)
 - Muy preocupado(a)
 - Algo preocupado(a)
 - No tan preocupado(a)
 - Nada preocupado(a)
 - No tiene suficiente conocimiento
- 19(a). ¿Cree usted que los resultados del Censo de los Estados Unidos ayudan más a un partido político (el Partido Republicano o el Partido Demócrata) que a otro, o cree que no es así?
- Sí
 - No
 - No tiene suficiente conocimiento
- 19(b). [*If R answered 19(a) with “Yes”.*] ¿En su opinión, qué partido se beneficiará más por los resultados del Censo de los Estados Unidos?
20. Por lo que usted sabe, ¿se utiliza el Censo para determinar si alguien está en este país legalmente, o no se utiliza así?
- Sí, se utiliza para determinar si alguien está en este país legalmente
 - No, no se utiliza para determinar si alguien está en este país legalmente
 - No tiene suficiente conocimiento
21. Por lo que usted sabe, ¿se utiliza el Censo para determinar el número de representantes que cada estado tendrá en el Congreso, o no se utiliza así?
- Sí, se utiliza para determinar el número de representantes que cada estado tendrá en el Congreso

- No, no se utiliza para determinar el número de representantes que cada estado tendrá en el Congreso
 - No tiene suficiente conocimiento
22. Por lo que usted sabe, ¿se utiliza el Censo para determinar la cantidad de dinero que las comunidades recibirán del gobierno, o no se utiliza así?
- Sí, se utiliza para determinar la cantidad de dinero que las comunidades recibirán del gobierno
 - No, no se utiliza para determinar la cantidad de dinero que las comunidades recibirán del gobierno
 - No tiene suficiente conocimiento
23. Por lo que usted sabe, ¿El Census Bureau (La Oficina del Censo) tiene que guardar la información personal que usted proporcionó en el formulario del Censo del 2020 de manera confidencial, o no tiene que hacerlo así?
- Sí, tiene que guardar la información personal proporcionada en el formulario del Censo del 2020 de manera confidencial
 - No, no tiene que guardar la información personal proporcionada en el formulario del Censo del 2020 de manera confidencial
 - No tiene suficiente conocimiento
24. *[If R identified any household member as being of Hispanic, Latino, or Spanish Origin (see Question 8).]* ¿Ha visto u oído algo recientemente de grupos cívicos, religiosos, medios de comunicación o grupos de la comunidad hispana/latina alentándole o desalentándole a llenar su formulario del Censo del año 2020?
- Sí
 - No

S2 Demographic and Balance Statistics

S2.1 Demographics

For race/ethnicity data, we relied on data provided by the vendor. We cross-checked their data with the responses to our survey, finding strong correspondence between reported race/ethnicity of the household and the vendor-provided race/ethnicity.

Table S2: Respondents' Race/Ethnicity and Partisanship by Survey Wave

	Wave 1	Wave 2
<u>Race/Ethnicity</u>		
Hispanic	55 (1.34%)	4,562 (92.52%)
African-American/Black	246 (5.99%)	3 (0.06%)
Asian-American/Asian	181 (4.41%)	1 (0.02%)
White	3,413 (83.16%)	13 (0.26%)
Native American/Inuit/Aleut	25 (0.61%)	1 (0.02%)
Native Hawaiian/Pacific Islander	10 (0.24%)	0 (0.00%)
Other	57 (1.39%)	2 (0.04%)
Not Provided	117 (2.85%)	349 (7.08%)
<u>Partisanship</u>		
Democrat	1,129 (27.51%)	1,526 (30.95%)
Republican	1,027 (25.02%)	653 (13.24%)
Independent	913 (22.25%)	758 (15.37%)
Other	75 (1.83%)	134 (2.72%)
Not Provided	960 (23.39%)	1,860 (37.72%)
<u>Country of Birth</u>		
<i>North America</i>		
All	2,916 (71.05%)	2,348 (47.62%)
United States	2,898 (70.61%)	2,346 (47.58%)
Canada	18 (0.44%)	2 (0.04%)
<i>Latin America</i>		
All	19 (0.46%)	621 (12.59%)
Mexico	1 (0.02%)	156 (3.16%)
Central America (Excl. Mexico)	1 (0.02%)	51 (1.03%)
Cuba	2 (0.05%)	74 (1.50%)
Puerto Rico	0 (0.00%)	159 (3.22%)
Caribbean (Excl. Cuba/Puerto Rico)	11 (0.27%)	52 (1.05%)
South America	4 (0.10%)	129 (2.62%)
<i>Europe</i>		
All	91 (2.22%)	83 (1.68%)
Northern Europe	21 (0.51%)	9 (0.18%)
Southern Europe	29 (0.71%)	62 (1.26%)
Eastern Europe	15 (0.37%)	6 (0.12%)
Western Europe	26 (0.63%)	6 (0.12%)
<i>Africa</i>		
All	18 (0.44%)	55 (1.12%)
Northern Africa	9 (0.22%)	27 (0.55%)

Continued on next page

Table S2 – *Continued from previous page*

	Wave 1	Wave 2
Middle Africa	7 (0.17%)	23 (0.47%)
Eastern Africa	0 (0.00%)	4 (0.08%)
Western Africa	2 (0.05%)	1 (0.02%)
<i>Asia</i>		
All	75 (1.83%)	44 (0.89%)
Central Asia	0 (0.00%)	1 (0.02%)
Southern Asia	22 (0.54%)	10 (0.20%)
South-Eastern Asia	18 (0.44%)	10 (0.20%)
Eastern Asia	28 (0.68%)	7 (0.14%)
Western Asia	7 (0.17%)	16 (0.32%)
<i>Oceania</i>		
All	1 (0.02%)	5 (0.10%)
Australia/New Zealand	1 (0.02%)	3 (0.06%)
Micronesia	0 (0.00%)	2 (0.04%)
Not Provided	984 (23.98%)	1,775 (36.00%)
Total	4104	4931

Table S2 shows the demographic and partisan composition of the first and second waves of our survey. Of these variables, the most important is race/ethnicity where 1.34% of the first wave of our survey was identified as “Hispanic” by Qualtrics which is substantially less than the percent identified as “Hispanic” in the second wave (92.52%). Other noticeable differences are found in the partisan breakdown in which 25.02% of the first wave of our survey were identified as Republicans which is 11.78 percentage points higher than the percent in second wave (13.24%). With that said, we found a reasonable number of respondents identified as Democrats. More specifically, Qualtrics identified 27.51% and 30.95% as being Democrats in the first and second waves of the survey, respectively. This is important since Hispanics tend to identify as being members of the Democratic party which is why we were concerned that this variable would be disproportionately represented in the second wave of our survey.

Table S3: Respondents’ Race/Ethnicity and Partisanship Compared to 2010 Census

	Survey	Census
<u>Race/Ethnicity</u>		
Hispanic	4,617 (51.10%)	50,477,594 (16.35%)
African-American/Black	249 (2.76%)	38,929,319 (12.61%)
Asian-American/Asian	182 (2.01%)	10,242,998 (3.32%)
White	3,426 (37.92%)	196,817,552 (63.75%)
Native American/Inuit/Aleut	26 (0.29%)	29,32,248 (0.95%)
Native Hawaiian/Pacific Islander	10 (0.11%)	540,013 (0.17%)

Continued on next page

Table S3 – Continued from previous page

	Survey	Census
Other	59 (0.65%)	8,488,805 (2.75%)
Not Provided	466 (5.16%)	317,009 (0.10%)
<u>Partisanship</u>		
Democrat	2,655 (29.39%)	123,251,219 (39.92%)
Republican	1,680 (18.59%)	72,524,327 (23.49%)
Independent	1,671 (18.49%)	96,328,608 (31.20%)
Other	209 (2.31%)	83,97,879 (2.72%)
Not Provided	2,820 (31.21%)	82,43,505 (2.67%)
<u>Country of Birth</u>		
<i>North America</i>		
All	5,264 (58.26%)	269,596,539 (87.32%)
United States	5,244 (58.04%)	268,789,539 (87.06%)
Canada	20 (0.22%)	807,000 (0.26%)
<i>Latin America</i>		
All	640 (7.08%)	21,224,000 (6.87%)
Mexico	157 (1.74%)	11,711,000 (3.79%)
Central America (Excl. Mexico)	52 (0.58%)	3,053,000 (0.99%)
Caribbean	298 (3.30%)	3,731,000 (1.21%)
South America	133 (1.47%)	2,730,000 (0.88%)
<i>Europe</i>		
All	174 (1.93%)	4,817,000 (1.56%)
<i>Africa</i>		
All	73 (0.81%)	1,607,000 (0.52%)
<i>Asia</i>		
All	119 (1.32%)	11,284,000 (3.65%)
<i>Oceania</i>		
All	6 (0.07%)	201,376 (0.07%)
Not Provided	2,759 (30.54%)	15,625 (0.01%)
Total	9035	308,745,538

Table S3 compares the demographic and partisan composition of both waves of our survey to the 2010 Census estimates. Since the U. S. Census Bureau does not ask about partisanship, we imputed the total number of Democrats, Republicans, and Independents using the 2010 population (308,745,538) and weighted party identification estimates from the 2008 American National Election Studies (ANES). For example, in the nationally weighted version of the 2008 ANES 39.92% of respondents said they identified with the Democratic Party. Multiplying that percentage by 308,745,538 yields an estimated 123,251,219 Democrats. We repeated this process for Republicans, Independents, and people who identified with another party (“Other”) or did not provide an

answer (“Not Provided”). The nationally weighted 2008 ANES were used to impute each of these categories.

Our sample is similar to the 2010 Census in some categories, but different in others. For example, in the 2010 Census 6.87% of the population was born in Latin American which is very close to the percent in our survey (7.08%). We find similar results for respondents Qualtrics identified as being born in Central American countries (excluding Mexico) which we define as Belize, Costa Rica, Dominican Republic, El Salvador, Guatemala, Honduras, Nicaragua, and Panama. In the 2010 Census, 0.99% of the population was born in these countries, whereas the same can be said for 0.58% of our respondents. There are a few other examples where our sample is similar to the 2010 Census (e.g., Asian-Americans, etc.), but our sample is mostly different which is why we employ post-stratification weights (see discussion in Section S3.1) before extrapolating our results to the 2020 Census.

S2.2 Balance Statistics

Balance statistics were calculated using the **MatchBalance** function of the **Matching** library in the **R** statistical software language. These are presented in Tables S4 and S5 for the citizenship question and Census prompt, respectively. In the second and third columns, we report the means for the treatment (\bar{X}_T) and control (\bar{X}_C) groups for each of the variables listed in the first column. Unadjusted and Bonferroni-corrected p -values are reported in the fourth (p) and fifth (\hat{p}) columns, respectively. We also estimated pairwise interactions for all variables, but none of these were statistically significant at the 0.05-level.

Table S4: Balance Statistics for Citizenship Question Treatment

Variable	$\bar{X}_{\text{Treatment}}$	\bar{X}_{Control}	p -value	\hat{p} -value
Wave	0.540	0.552	0.266	1.000
<u>Race/Ethnicity</u>				
Hispanic	0.508	0.514	0.584	1.000
African-American/Black	0.026	0.029	0.526	1.000
Asian-American/Asian	0.022	0.019	0.267	1.000
White	0.382	0.376	0.550	1.000
Native American/Inuit/Aleut	0.003	0.003	0.712	1.000
Native Hawaiian/Pacific Islander	0.001	0.001	0.518	1.000
Other	0.022	0.025	0.325	1.000
Not Provided	0.050	0.053	0.638	1.000
<u>Party</u>				
Democrat	0.294	0.294	0.981	1.000
Republican	0.187	0.185	0.753	1.000
Independent	0.183	0.187	0.562	1.000
Other	0.022	0.025	0.325	1.000

Continued on next page

Table S4 – *Continued from previous page*

Variable	$\bar{X}_{\text{Treatment}}$	\bar{X}_{Control}	<i>p</i> -value	\hat{p} -value
Not Provided	0.315	0.310	0.605	1.000
Country of Birth				
<i>North America</i>				
Canada	0.002	0.002	0.669	1.000
United States	0.579	0.582	0.795	1.000
<i>Latin America</i>				
Mexico	0.017	0.017	0.982	1.000
Central America (Excl. Mexico)	0.004	0.007	0.055	1.000
Cuba	0.009	0.008	0.787	1.000
Puerto Rico	0.020	0.015	0.082	1.000
Caribbean (Excl. Cuba/Puerto Rico)	0.007	0.007	0.928	1.000
South America	0.014	0.015	0.576	1.000
<i>Europe</i>				
Northern Europe	0.004	0.002	0.064	1.000
Southern Europe	0.010	0.010	0.785	1.000
Eastern Europe	0.002	0.003	0.283	1.000
Western Europe	0.003	0.004	0.164	1.000
<i>Africa</i>				
Northern Africa	0.003	0.005	0.329	1.000
Middle Africa	0.003	0.003	0.980	1.000
Eastern African	0.001	>0.001	0.314	1.000
Western Africa	>0.001	>0.001	0.568	1.000
<i>Asia</i>				
Central Asia	>0.001	0.000	0.317	1.000
Southern Asia	0.004	0.003	0.277	1.000
South-Eastern Asia	0.003	0.003	0.723	1.000
Eastern Asia	0.003	0.004	0.412	1.000
Western Asia	0.003	0.002	0.818	1.000
<i>Oceania</i>				
Australia/New Zealand	>0.001	>0.001	0.993	1.000
Micronesia	>0.001	>0.001	0.995	1.000
Not Provided	0.308	0.303	0.656	1.000

Note: Unadjusted and Bonferroni-corrected *p*-values from two-sample *t*-tests reported in the last two columns. No pairwise interactions were statistically significant at the 0.05-level.

To calculate balance statistics for the country of birth, we grouped country into regions as defined by the World Bank Development Indicators which were obtained from the **countrycode** library in the **R** statistical software language. The only changes made to the World Bank regions

were in North America where we separated the United States from Canada and used Latin American subdivisions that were more consistent with our survey. Ultimately, our respondents were born in 100 different countries which were organized into 24 different regions.

Table S5: Balance Statistics for Census Prompt Treatment

Variable	$\bar{X}_{\text{Treatment}}$	\bar{X}_{Control}	p -value	\hat{p} -value
Wave	0.545	0.546	0.904	1.000
<u>Race/Ethnicity</u>				
Hispanic	0.510	0.512	0.898	1.000
African-American/Black	0.026	0.029	0.260	1.000
Asian-American/Asian	0.019	0.021	0.479	1.000
White	0.383	0.376	0.486	1.000
Native American/Inuit/Aleut	0.002	0.003	0.474	1.000
Native Hawaiian/Pacific Islander	0.001	0.001	0.965	1.000
Other	0.023	0.024	0.776	1.000
Not Provided	0.053	0.050	0.551	1.000
<u>Party</u>				
Democrat	0.295	0.293	0.848	1.000
Republican	0.190	0.182	0.309	1.000
Independent	0.189	0.181	0.297	1.000
Other	0.023	0.024	0.776	1.000
Not Provided	0.303	0.321	0.068	1.000
<u>Country of Birth</u>				
<i>North America</i>				
Canada	0.002	0.003	0.199	1.000
United States	0.576	0.584	0.439	1.000
<i>Latin America</i>				
Mexico	0.018	0.017	0.796	1.000
Central America (Excl. Mexico)	0.005	0.007	0.311	1.000
Cuba	0.008	0.009	0.915	1.000
Puerto Rico	0.019	0.017	0.460	1.000
Caribbean (Excl. Cuba/Puerto Rico)	0.007	0.007	0.789	1.000
South America	0.013	0.016	0.251	1.000
<i>Europe</i>				
Northern Europe	0.004	0.003	0.420	1.000
Southern Europe	0.009	0.011	0.415	1.000
Eastern Europe	0.003	0.002	0.249	1.000
Western Europe	0.004	0.003	0.665	1.000
<i>Africa</i>				
Northern Africa	0.004	0.004	0.676	1.000

Continued on next page

Table S5 – *Continued from previous page*

Variable	$\bar{X}_{\text{Treatment}}$	\bar{X}_{Control}	p -value	\hat{p} -value
Middle Africa	0.004	0.003	0.242	1.000
Eastern African	>0.001	0.001	0.328	1.000
Western Africa	>0.001	>0.001	0.579	1.000
<i>Asia</i>				
Central Asia	>0.001	0.000	0.317	1.000
Southern Asia	0.004	0.003	0.432	1.000
South-Eastern Asia	0.003	0.003	0.651	1.000
Eastern Asia	0.005	0.003	0.206	1.000
Western Asia	0.002	0.003	0.576	1.000
<i>Oceania</i>				
Australia/New Zealand	0.000	0.001	0.045	1.000
Micronesia	>0.001	0.000	0.157	1.000
Not Provided	0.309	0.302	0.468	1.000

Note: Unadjusted and Bonferroni-corrected p -values from two-sample t -tests reported in the last two columns. No pairwise interactions were statistically significant at the 0.05-level.

Beginning with Table S4, all Bonferroni-corrected p -values are well above the 0.05 threshold and approximate 1. Not only does this demonstrate we have reasonable balance across the treatment and control groups for our Citizenship Treatment, but none of the unadjusted p -values are below 0.05 which gives us additional confidence that our sample is equally distributed across both conditions. A similar results is found in Table S5. Again, the Bonferroni-corrected p -values approach 1 and only one unadjusted p -value is below the 0.05-level (respondents from Australia/New Zealand). This demonstrates the characteristics of the respondents who received the Census prompt are essentially the same as those who did not receive this treatment.

S3 Additional Analyses

S3.1 Survey Weights

In order to make our sample more nationally representative, we created post-stratification weights using the **rake** function from the **survey** library in the **R** statistical software language. Generally speaking, a raking algorithm take known population distributions and creates sample weights in order to make the sample’s marginal distributions identical to their counterparts in the population. The process is iterative, meaning initial weights are created to make the marginal distribution of the first variable identical in the sample and population. Those weights are then adjusted so the marginal distributions of the second variable matches the population distribution. So forth and so on, until the algorithm converges and you have a weighted sample in which the marginal distributions of the variables you provide are identical to those in the population.

Our vendor was able to provide the following information for our respondents: Race/Ethnicity, Country of Birth, Party Identification, Zip Code, and Religion. Of these variables, we had the best coverage for Race/Ethnicity and Zip Code which was provided for 94.99% and 95.03% of our respondents, respectively. From there, the information provided by Qualtrics becomes increasingly sparse. Party identification was not provided for 31.21% of our respondents. Qualtrics did not identify the country of birth for 22.39% of our respondents. Not only was religion not provided for 43.00% of our respondents, but we also could not find population distributions for all 20 religions our vendor provided.

Given these limitations, we created post-stratification weights using the following variables: (1) Race/Ethnicity, (2) Zip Code, and (3) whether the individual identified as a Democrat. The raking algorithm would not converge in a reasonable amount of time using the other configurations which is why we focused our efforts on these three variables. Using our weighted sample, we then estimated the effect of receiving the citizenship question on the percentage of the household reported as being of Hispanic, Latino, or Spanish origin using the `svytest` function from the `survey` library mentioned above. When this was done, we still found the citizenship question lead to a significant underreporting (1.97 percentage points) of Hispanic household members (t -statistic = 3.205, p -value = 0.001) which gives some evidence that the results from our experiment have some generalizability.

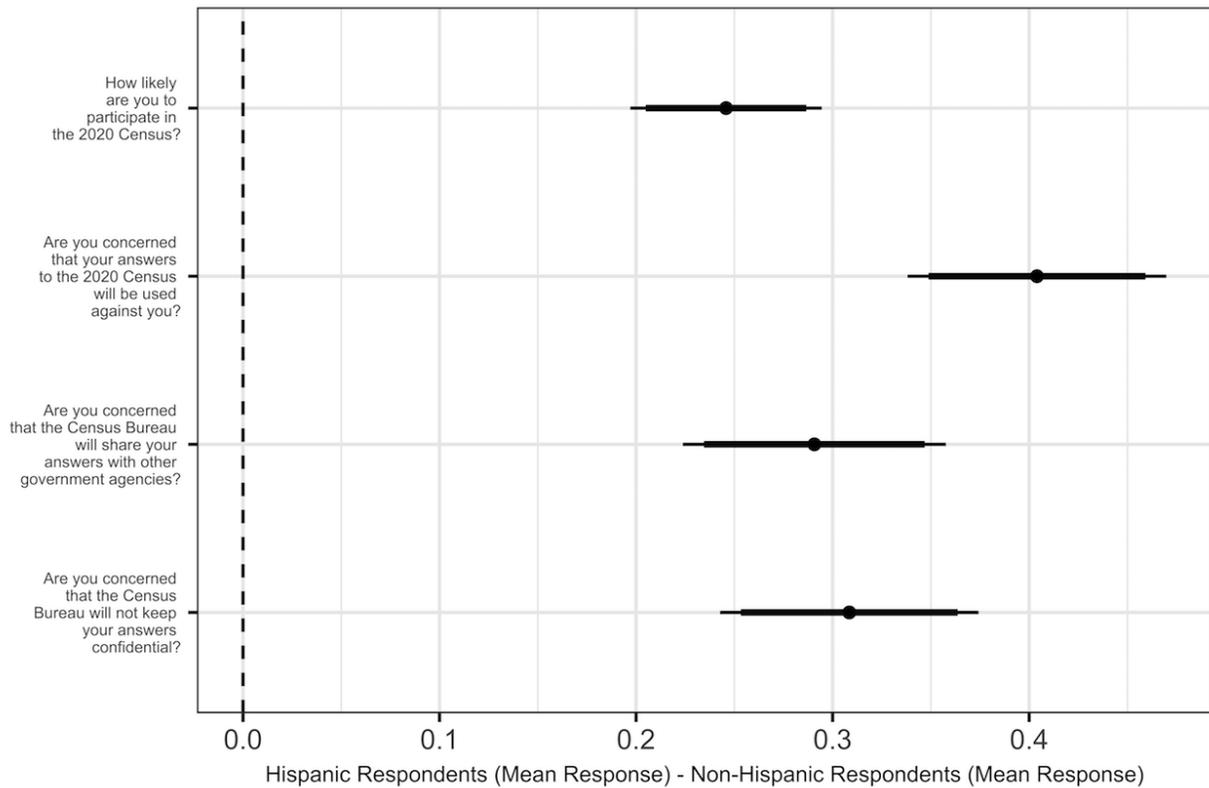
To extrapolate to the 2020 Census we used the `svymean` function from the `survey` library to derive weighted estimates (and their standard errors) of the proportion of the household that is Hispanic under treatment (11.85) and control (13.82). We then multiplied the total population according to the 2010 Census (308,745,538) by these weighted estimates which yielded an estimated 36,587,056 and 42,659,124 Hispanics, respectively. We then subtracted the treatment estimate (36,587,056) from the control estimate (42,659,124) to arrive at 6,072,068 fewer Hispanics reported.

S3.2 Census Opinion Questions

At the end of the survey, we asked our respondents 13 questions (or 14 questions when Hispanic household members were listed) about their opinions of the 2020 Census and the Census Bureau more broadly. Generally speaking, respondents skipped a large number of these questions which means the respondents who did offer opinions were likely different from those who did not. This caveat aside, we found Hispanics were generally more concerned about the 2020 Census.

Fig. S1 shows Hispanic respondents generally are (1) less likely to participate in the 2020 Census and more concerned (2) their 2020 Census answers will be used against them, (3) the Census Bureau will share their answers with other government agencies, and (4) their answers will not be kept confidential. We think these additional results clearly demonstrate Hispanics are a “hard to count” population meaning any changes to the Census that could disproportionately affect this population should be made cautiously.

Figure S1: Hispanic Respondents Are Generally More Concerned About the 2020 Census



Note: Difference in the average response for Hispanic and non-Hispanic respondents. All questions are scaled so higher values mean a more negative answer. For example, “How likely are you to participate in the 2020 Census?” is coded as 1 = “Definitely will” and 5 = “Definitely will not.” Similarly, “Are you concerned that the Census Bureau will not keep your answers confidential?” is coded as 1 = “Extremely concerned” and 5 = “Not at all concerned.” The full questions are reported in Section S1.4 of the Supporting Material (SI). Thicker (—) and thinner (—) lines represent 90 and 95-percent confidence intervals, respectively.

S3.3 Additional Hypotheses

In our pre-analysis plan, we registered several hypotheses and expectations. In this section, we will discuss some of the hypotheses that we did *not* discuss in the main text.

Survey respondents who receive the Census prompt are less likely to begin the survey.

In our pre-registration we hypothesized that the invitees who receive the Census prompt will begin the survey at a rate of 4%, but we found this to be intractable given some unforeseen limitations to our survey. More specifically, we were unable to determine who received the link to the survey, meaning we only have data on those who actually started the survey. This makes it difficult, if not impossible, to directly test this hypothesis.

To gain some traction, we calculated the proportion of early questions skipped by respondents. Here, we define Questions 1, 2, and 3 as “early” questions. Respondents receiving the Census prompt skipped 8.68 percent of these questions, whereas those not receiving the Census prompt skipped 7.33 percent. This 1.35 percentage point difference is statistically significant at the 0.02-level (t -statistic = 2.523, p -value = 0.012).

We also conducted an additional analysis on the number of reported household members which was the first question asked in the survey. Respondents who receive the Census Prompt, however, report smaller household sizes – an average household size of 2.75 under the treatment condition compared to 2.83 under the control condition, a significant drop (t -statistic = 2.07, p -value = 0.038). The drop is somewhat larger among Hispanics (3.04 under treatment vs. 3.15 under the control), although this difference is only statistically significant at the 0.07-level (t -statistic = 1.826, p -value = 0.068). We find a comparable, albeit smaller and insignificant, pattern among non-Hispanics (2.45 under the treatment vs. 2.50 for the control group; t -statistic = 1.028, p -value = 0.304).

Survey respondents who receive the citizenship question will answer fewer questions.

In the main text, we operationalize this hypothesis using the percent of questions skipped, but in our pre-registration we suggested those who receive the citizenship question should complete the survey at a rate of 80%. We tested this hypothesis by creating a dummy variable which equals 1 when respondents skipped more than 80 percent of the questions. Using this variable, we found 18.59 percent of respondents who received the citizenship question skipped more than 80 percent of the questions. Of those who did not receive the citizenship question, 14.94 percent skipped more than 80 percent of the questions. The 3.65 percentage point difference is also statistically significant ($\chi^2 = 9.788$, p -value = 0.002).

We also conducted additional analyses regarding questions about respondent’s age and date-of-birth. Conditional on the number of household members initially reported, respondents who received the Citizenship Treatment were significantly more likely to skip the questions concerning household members’ ages, on average by 3.32 percentage points (t -statistic = 4.111, p -value less than 0.001). The effect is stronger among Hispanics, who experience an increase in questions skipped of 4.56 percentage points (t -statistic = 3.597, p -value less than 0.001).

We again find large, significant effects among Hispanic respondents who report being born in either Mexico or a Central American country, where receiving the Citizenship Treatment increases the percent of questions skipped on household members' ages by 10.95 percentage points (t -statistic = 3.274, p -value = 0.001). The corresponding increase among Hispanic respondents listing Cuba or Puerto Rico as their birth country is 0.14 (t -statistic = 0.045, p -value = 0.964). We again find a smaller 2.27 percentage point effect among non-Hispanics (t -statistic = 2.423, p -value = 0.015).

Those who receive both the citizenship question and census prompt should show more pronounced effects.

In order to test this hypothesis, we estimated a simple linear regression in which the proportion of questions skipped was regressed on the interaction between our Citizenship and Census Prompt Treatment. Ultimately, the interaction was not statistically significant at the 0.05-level (t -statistic = 1.392, p -value = 0.164) which we expect may be due to a number of factors.

First, we were unable to re-contact respondents who dropped out after receiving the Census Prompt, but before receiving the Citizenship Treatment. Consequently, of those who received the Census Prompt, we only observe the effect of the Citizenship Treatment for those who responded the *least* to the Census Prompt.

Second, the Census Prompt is likely underpowered. The concern from the Census Bureau is that individuals will receive either an email or envelope that says "United States Census Bureau." That is quite a bit different from our treatment which simply added a checkbox at the bottom of a 2-page description of how we will protect their confidentiality. Those confidentiality assurances and the Institutional Review Board (IRB) likely altered or diminished the effect of the checkbox we introduced.

Third, the effect of the Census Prompt is likely more pronounced on unit non-response. Individuals see an email or envelope that says "United States Census Bureau" and simply do not respond. Our survey is designed to test item non-response, but we cannot effectively measure unit non-response since we do not know who received the survey link and decided not to participate.

Finally, our respondents are not only paid to complete the survey, but they are part of a Qualtrics panel. Both factors give individuals strong incentives to complete the survey and given their repeated exposure to survey instruments they are also less likely to be concerned that they would see any detrimental effects from their participation. Consequently, when we say we will protect their confidentiality they likely have some confidences in those assurances, otherwise they would not be members of a Qualtrics panel.

We anticipate that respondents identifying as Democrat/Leaning Democratic are more likely to respond to either or both treatments negatively than are those who identify as Republican/Leans Republican.

We found some evidence of partisan differences. Democrats who received the citizenship question skipped 25.16 percent of the questions which is 3.14 percentage points more than the percent of

questions skipped by those those who did not receive the citizenship question (22.02). Hispanics who were identify as Democrats by Qualtrics skipped 5.64 percentage points more questions when they received the citizenship question (t -statistic = 2.808, p -value = 0.005). No significant effects were found for Hispanics who identified as Republican (t -statistic = 1.114, p -value = 0.266) or Independent (t -statistic = 0.203, p -value = 0.839).

We anticipate little or no treatment effect among Puerto Ricans and Cuban Americans, but a negative treatment effect for both treatments among Mexicans and Central Americans.

Among the respondents from Mexico and Central America ($n = 240$), the citizenship question does seem to affect their response rate. More specifically, Hispanics who Qualtrics identified as being from either Mexico or Central American skipped 9.93 percent of the questions when they did not receive the citizenship question. When the citizenship question was randomly assigned this percentage increased to 20.97 percent and this 11.04 percentage point difference was statistically significant at the 0.05-level (t -statistic = 3.298, p -value = 0.001). When Hispanics who Qualtrics identifies as originating from Puerto Rico or Cuba ($n = 235$) received the citizenship question they skipped 13.56 percent of the survey. Under the control condition, these same respondents skipped 11.78 percent of the survey. This insignificant difference (t -statistic = 0.566, p -value = 0.572) and the significant difference for Hispanic respondents originating from Mexico and Central American provides evidence consistent with our pre-registered hypothesis.

We anticipate that Latina/o and non-whites will respond differently to the attitude questions we posed than non-Latina/o and whites, regardless of treatment status.

This hypothesis is discussed in Section S3.2 of the SI.

We anticipate respondents receiving either or both treatment conditions, relative to the control conditions will have different attitudes towards the Census.

We did not find evidence consistent with hypothesis. When we asked “How likely are you to participate in the 2020 United States Census? By participate, we mean fill out and mail in a Census form or fill one out online?” there were 5 response options (other than “Don’t Know”) ranging from “Definitely will” (1) to “Definitely will not” (5).

For this question we found no statistically significant difference between the mean responses for those who did (1.73) and did not (1.74) receive the citizenship question (t -statistic = 0.390, p -value = 0.697). The same can be said for the Census prompt. Although the effect is more pronounced, we again found no significant difference between the mean responses for the treatment (1.71) and control (1.75) groups (t -statistic = 1.890, p -value = 0.060).

We found the same results when we asked “How concerned are you, if at all, that the answers you provide to the 2020 Census will be used against you?” For this question there were again 5 response options (other than “Don’t Know”) ranging from “Extremely concerned (1)” to “Not at all concerned” (5). To make this variable comparable to the other questions, we inverted the scale so higher values implied greater concern.

Again, we found no statistically significant difference between the mean responses for those who did (2.22) and did not (2.24) receive the citizenship question (t -statistic = 0.700, p -value = 0.484). The same can be said for the Census prompt. Although the effect is more pronounced, we again found no significant difference between the mean responses for the treatment (2.21) and control (2.25) groups (t -statistic = 1.257, p -value = 0.209).

When we asked “How concerned are you, if at all, that the Census Bureau will share answers to the 2020 Census with other government agencies?” we found the same results. For this question the 5 response options (other than “Don’t Know”) were again inverted so “Extremely concerned” was re-coded as 5 and “Not at all concerned” was re-coded as 1. This was done to make this question comparable to the rest with higher values implying more concern.

We again found no statistically significant difference between the mean responses for those who did (2.61) and did not (2.64) receive the citizenship question (t -statistic = 1.114, p -value = 0.265). The same can be said for the Census prompt where no significant difference between the mean responses for the treatment (2.60) and control (2.65) groups (t -statistic = 1.220, p -value = 0.223).

Finally, neither treatment significantly affected responses to the following question “How concerned are you, if at all, that the Census Bureau will not keep answers to the 2020 Census confidential?” The 5 response options (other than “Don’t Know”) were again inverted so “Extremely concerned” was re-coded as 5 and “Not at all concerned” was re-coded as 1.

For this question we found no statistically significant difference between the mean responses for those who did (2.70) and did not (2.70) receive the citizenship question (t -statistic = 0.087, p -value = 0.931). The same can be said for the Census prompt. We again found no significant difference between the mean responses for the treatment (2.71) and control (2.69) groups (t -statistic = 0.748, p -value = 0.454).

Although these results are not consistent with our pre-registered expectations, the non-response rate in this section is noticeably higher than the rest. For example, respondents skipped 36.58% of the questions regarding their opinions towards the Census which is much higher than the 8.00% of questions skipped in the first part of our survey. Since we do not know the opinions of these missing respondents, it is difficult to say how this affects our original hypothesis, but we can say that the respondents who are answering these questions are likely different from those who did not.

S3.4 Congressional District Analysis

Since Hispanic populations are unevenly distributed across the United States, we were interested in whether certain congressional districts will be disproportionately affected by introducing the citizenship question. Although we have at least 1 respondent in all 435 congressional districts, our sample is not balanced across all districts which makes it difficult to properly estimate the effect of receiving the citizenship question within a single congressional district. One way to achieve this end is to subset our data by congressional district and then re-estimate the effect of receiving the citizenship question on the percent of the household reported as being of Hispanic, Latino, or Spanish Origin. We did this for all 435 congressional districts and identified the 10 districts in which our Citizenship Treatment had the most pronounced effect.

Table S6: Congressional Districts with Largest Marginal Effect (Citizenship Question)

District	Treatment (N)	Control (N)	Treatment (Mean)	Control (Mean)	Diff.	<i>p</i> -value
CA-22	9	13	25.00	78.57	-53.57	0.01
CA-48	7	3	35.71	88.89	-53.17	0.11
CA-37	8	6	2.50	54.17	-51.67	0.01
CA-1	6	6	12.50	63.89	-51.39	0.03
CA-4	7	10	20.00	70.00	-50.00	0.04
OK-4	2	8	16.67	62.50	-45.83	0.27
CA-39	9	7	22.22	66.67	-44.44	0.05
CA-38	10	7	46.07	88.10	-42.02	0.03
TX-35	10	13	43.33	84.62	-41.28	0.03
CA-45	5	11	15.00	53.03	-38.03	0.08

Note: This table shows the congressional districts in which the respondents seemed to respond the most to our Citizenship Treatment.

Table S6 reports these initial results. Not only do districts from California represent 8 of the 10 districts most affected by our Citizenship Treatment, but many of the statistically significant differences exist within Southern California. Using an unadjusted *p*-value from a two-sample *t*-test, we find that California’s 1st, 4th, 22nd, 37th, 38th, and 39th districts all show significant ($p < 0.05$) declines in the percentage of household members reported as being Hispanic when respondents in those districts received the citizenship question. A similar decline is found in California’s 45th district, but it is only statistically significant at the 0.10-level. All of these districts except for California’s 1st, 4th and 22nd are in the Los Angeles metropolitan area and California’s 22nd district is around 176 miles (or a 3 hour drive) away.

S3.5 Imputation Analysis

Given that the U.S. Census Bureau often imputes missing information from survey respondents, we replicate our results regarding the percent of the household reported as being of Hispanic, Latino, or Spanish origin using predictive means matching. We could not find any information on the algorithm used by the U. S. Census Bureau so we chose an imputation algorithm that is commonly used in the social sciences.

In predictive mean matching (PMM), respondents’ missing values are imputed with random values from other respondents whose regression-predicted values are closest to the respondents whose values are being imputed. The main benefit of PMM is that the imputed values are plausible which is why we chose this method for our replication exercise. The PMM algorithm was implemented using the **mice** function from the **mice** library of the **R** statistical software language. This package uses “Multiple Imputation by Chained Equations” which applies the imputation algorithm several times, then the aggregated results are used for the final imputed values. For our study, we used the default **mice** settings which is 5 iterations of 5 imputations – producing 25 different versions of the imputed data. Missing values were then replaced using the **complete**

function from the same library.

Given that survey responses are influenced by whether respondents were randomly assigned to either the treatment or control conditions, we subdivided our data into those who did and did not receive the citizenship question. Missing household member demographic information was then imputed. Once this was done, we then conducted a *t*-test using the imputed data from the treatment and control data.

Since household size varies by respondent, we used an iterative imputation process beginning with “Person 1.” The steps are as follows:

1. Using PMM impute Person 1’s NAs using all available demographic information (age, gender, race, hispanic origin) for Person 1. Replace Person 1’s NAs with imputed values.
2. For respondents who listed 2 or more household members, use the updated Person 1 values (original and imputed) and all available demographic information (age, gender, race, hispanic origin) for Person 2 to impute Person 2’s NAs with PMM. Replace Person 2’s NAs with imputed values.
3. For respondents who listed 3 or more household members, use the updated Person 1 values (original and imputed), updated Person 2 values (original and imputed), and all available demographic information (age, gender, race, hispanic origin) for Person 3 to impute Person 3’s NAs with PMM. Replace Person 3’s NAs with imputed values.
4. Repeat these steps until all missing demographic information is imputed for all household members reported by the respondent.

Using the imputed data, we found those receiving the Citizenship Treatment reported that a smaller share (40.75) of their household members are of Hispanic origin compared to those in the treatment condition (45.08). This difference is still statistically significant at the 0.0001-level (*t*-statistic = 4.318, *p*-value less than 0.0001) which suggests imputing the missing data does not diminish the marginal treatment effect we report in the main text. Using the imputed data, Hispanic respondents receiving the Citizenship Treatment reported 64.45 percent of their household members were of Hispanic origin compared to 73.02 percent in the control condition. This 8.57 percentage point difference is statistically significant at the 0.0001-level (*t*-statistic = 6.891, *p*-value less than 0.0001) which is consistent with our main analysis. The same cannot be said for Non-Hispanic respondents. Using the imputed data, when these respondents received the Citizenship Treatment they reported 16.97 percent of their household members were Hispanic as compared to 16.23 percent in the control condition. This 0.74 percentage point difference is not statistically significant (*t*-statistic = 0.682, *p*-value = 0.495).

S3.6 Direct Costs Analysis

In this subsection, we try to extrapolate the costs of undercounting 6,072,068 Hispanics in the 2020 Census. It is difficult to say precisely how much it would cost, but we do our best using 2000 and 2010 Census estimates. At the high-end, re-contacting these households could cost anywhere

between \$1,088,515,183 to \$1,287,212,181 in projected dollars which is a large percentage of the approximately \$5 billion Census operating budget. The Census Bureau spent \$1,589,397,886 following up with 47,235,198 households in 2010 which equates to a per-household rate of \$33.65 for the first followup. Households that had to be recontacted two and three times cost the Census Bureau \$84.09, and \$142.53, respectively. If we assume the 6,072,068 fewer Hispanics all live in unique households, then introducing the citizenship question could cost the Census Bureau up to \$204,325,088 in order to fill in the missing information. If these same households need to be followed-up with two and three times, then the citizenship question could cost the Census Bureau \$510,600,198 and \$865,451,852, respectively. Using the first followup rate (one additional contact at \$33.64) and the 95-percent confidence interval we report in the main text (5,761,284 to 6,382,820), filling in the missing information generated from introducing the citizenship question could cost anywhere between \$193,867,207 to \$214,781,893 assuming each missing Hispanic lives in a unique household. If we use the third followup rate (three additional contacts equaling \$142.53 for a single household), contacting these missing Hispanic household members could cost anywhere between \$821,155,809 to \$909,743,335 in 2010 dollars.

Between the 2000 and 2010 Census, the rate for conducting a single household followup increased from \$26.58 to \$33.65 which represents a 26.59 percent increase over the 2000 Census first followup rate. The 2010 Census second and third followup rates also increased by 90.20 and 96.32 percent based on the costs for the 2000 Census. If we assume the 2020 Census will see similar rate increases, then we can predict the first, second, and third followups will cost \$42.60, \$159.94, and \$279.81 per household, respectively. Using the first predicted followup rate for the 2020 Census (one additional contact at \$42.60) and the estimated 95-percent confidence interval we report in the main text (5,761,284 to 6,382,820), filling in the missing information generated by the citizenship question could cost anywhere between \$245,430,698 to \$271,908,132 assuming each missing Hispanic lives in a unique household. If we use the third predicted followup rate for the 2020 Census (three additional contacts equaling \$279.81 for a single household), contacting these missing Hispanic household members could cost anywhere between \$1,612,064,876 to \$1,785,976,864 in projected dollars.

S3.7 Modeling Proportions

Since our main dependent variables are proportions, a censored regression is more appropriate for estimating our marginal treatment effects. In this subsection, we present the results from censored regressions in which the model restricts the distribution of the dependent variable between 0 and 1. All models were estimated using the **censReg** function from the **censReg** library in the **R** statistical software language and Table S7 reports the results. We find our main results hold when the appropriate distributional assumptions are made regarding our dependent variables. Similar results are found when beta regressions are estimated.

Table S7: Re-Estimating Our Marginal Treatment Effects Using Censored Regressions

	<i>Dependent variable:</i>			
	Skipped Questions		Hispanic Household Members	
	(1)	(2)	(3)	(4)
Constant	0.243*** (0.006)	0.262*** (0.006)	-0.550*** (0.056)	-0.709*** (0.059)
Citizenship Question	0.040*** (0.009)		-0.337*** (0.072)	
Census Prompt		0.001 (0.009)		-0.013 (0.071)
$\log(\sigma)$	-0.921*** (0.008)	-0.920*** (0.008)	0.991*** (0.028)	0.993*** (0.028)
N	8,475	8,475	8,475	8,475
Log Lik	-5,024.215	-5,034.741	-7,809.960	-7,821.139
AIC	10,054.430	10,075.480	15,625.920	15,648.280

Note: “Skipped Questions” is the percentage of questions skipped. Higher values imply the respondent skipped more questions. “Hispanic Household Members” is the percentage of household members the respondent identified as being of Hispanic, Latino, or Spanish origin. Standard errors are reported in parentheses. Levels of significance are as follows: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$