A Valid Analysis of a Small Subsample: The Case of Non-Citizen Registration and Voting

Jesse Richman

Old Dominion University

jrichman@odu.edu

David C. Earnest

Old Dominion University, and

Gulshan Chattha

George Mason University

Abstract.

The development of large sample surveys creates new opportunities for analysis of sub-populations that would hitherto have been impossible to examine systematically. But it also raises key challenges. Low level measurement error can potentially lead to substantial biases in estimates drawn from small subsamples. This study details strategies researchers may take to make inferences in the context of this subsample-response-error problem. In the non-citizen voting case, which recently has received substantial attention, we show that attention to any of these strategies -- group-specific response error estimates, correlated higher-frequency events, test-retest validity, or analysis of associated hypotheses – produces significant evidence that non-citizens participated in recent US elections. It is also important that researchers aiming to debunk a study using this argument have a test with sufficient power. The analysis reaffirms the validity of the core claim made by Richman, Chattha, and Earnest (2014): a small percentage of non-citizens vote in US elections.

Ansolabehere, Luks, and Shaffner (2015) issued a methodological caution concerning work that aims to use small subsets of large survey datasets to make inferences about sub-populations of interest: error in the identification of subpopulation members may bias measurements. Despite the potential value of this argument, their effort to apply this caution to dismiss or 'debunk' the Richman et. al. (2014) study of non-citizen voting falls short for several reasons. These reasons fall into three broad categories: a lack of statistical power, problems with the assumptions or hypotheses needed to maintain their critique, and problems with the conclusions they draw from the critique itself.

Ansolabehere et. al. (2015) failed to consider key alternative theories that arguably better explain the patterns they identified, and their hypotheses concerning response error do not fit with patterns in the data. Furthermore, even if their arguments about response error are taken at face value, the CCES survey continues to provide substantial evidence that non-citizens participate in U.S. elections that they ignored in their paper. Thence even if there is some validity to a part of their argument (a few citizens may have misstated their citizenship status and thereby biased the Richman et. al. (2014) estimates) they go far too far when claiming that "the likely percent of non-citizen voters in recent US elections is 0."

The claim that there is no non-citizen participation made by Ansolabehere et. al. (2015) is particularly striking in light of a variety of other evidence which would seem to strongly suggest at least some non-citizen participation in US elections. For instance, in 2014 North Carolina officials reported that nearly one percent of so called 'dreamers' (undocumented non-citizens brought to the US as children) were registered to vote (Richman 2016a). 2016 reports from some Virginia counties found that a substantial number of non-citizens had been removed from voter rolls for that reason -- between 0.3 and 4.8 percent of the county non-citizen population (Richman 2016b). A simple Google search also highlights newspaper articles and internet help forums concerning the plight of particular non-citizens who registered to vote and faced or fear legal consequences.[1] The "National Hispanic Survey" in 2013 surveyed a sample that appeared to consist of more than half non-citizens (hence the response error issues that Ansolabehere et. al. highlight present a much smaller threat to validity) but found that 13 percent of non-citizens were registered to vote (McLaughlin 2013). Nonetheless, Schaffner (2016) writes that "there is no evidence that non-citizens have voted in recent U.S. elections."[2]

Subpopulations and Subsamples

A challenge for any research design focused on understanding the behavior of a small group within a broader population is accurate identification of members of the group for study.

---

[1] See for instance: http://www.immihelp.com/forum/showthread.php/163789-Permanent-Resident-Registered-to-Vote
http://www.nytimes.com/2010/10/17/nyregion/17voting.html?_r=0
http://www.usavisacounsel.com/articles/i-am-not-a-u-s-citizen-but-i-registered-to-vote%E2%80%A6-and-even-voted.htm

[2] This is an assertion that holds up only if (a) North Carolina, Virginia and other states accurately purge every non-citizen from their voting roles; or (b) none of the registered non-citizens actually vote. Hence the claim of "no evidence" relies strong assumptions that are unlikely.

Non-citizens make up a small portion of the overall US voting-age population and self-reported non-citizens make up a small portion of the typical CCES sample. This raises potential risks for inferences about the behavior of non-citizens, and these risks are most extreme when the behavior being analyzed is one much more common among citizens than non-citizens such as voting. Consequently, there is a risk that inferences will be biased by erroneously mis-assigned individuals who are not part of the target group but get misidentified as group members.

Because of this challenge, Richman et. al. (2014) contained an appendix with multiple analyses aimed at validating citizen-status self-reports, including the racial demographics, geographic distribution, and issue attitudes of non-citizen respondents. Without ever addressing or acknowledging the multiple validation approaches taken in the Richman et. al. (2014) appendix, Ansolabehere et. al. (2015) argue that the results of the Richman et. al. (2014) study on non-citizen participation "are completely accounted for by very low frequency measurement error."

Because of the possibility that measurement error could bias their results, authors of studies utilizing subsamples of large national surveys should undertake a careful analysis of the characteristics of the subsample and the nature of response error in order to quantify the magnitude of potential biases, and evaluate whether their results can be accounted for by measurement error. The appendix of Richman et. al. (2014) provides precisely such a careful analysis of this risk. In this study, we go farther as data released since the earlier papers were published are now available.

This response to Ansolabehere, Luks, and Schaffner's critique of Richman et. al. (2014) has three sections. After we first point out that their tests lacked statistical power, the second section presents evidence that the citizen status variable in the CCES is more accurate than Ansolabehere et. al. (2015) claim it is, with much of the error accounted for by intentional or unintentional errors made by non-citizens who claim to be citizens. This undermines their claim that response error debunks the Richman et. al. (2014) result. The third section sets aside the evidence from the first section and assumes that Ansolabehere et. al. (2015) were in fact correct about response error. We show that even if their response error argument is correct, there is still significant evidence of non-citizen participation in the U.S. electoral system.

All of the approaches to assessing the validity of inferences made from a subsample produce results counter to the claim made by Ansolabehere et. al. (2015) that "the likely percent of non-citizen voters in recent US elections is 0." While we have always recognized that some response error by citizens may potentially have biased our results, the evidence presented shows that this error is far too small to support the claim Schaffner (2016) made of having "debunked" the Richman et. al. (2014) study.

1. **An Under Powered Test**

A fundamental flaw with the Ansolabehere et. al. (2015) critique is that it lacks statistical power. The primary analysis is of validated voting in the 2010 midterm election. In the 2010 midterm CCES cross-sectional file 7 non-citizens cast validated votes. Excluding Virginia (no record checks were possible because of state law) this implies that 1.3 percent of non-citizens

cast validated votes in 2010. Ansolabehere et. al. examine verified voting in 2010 by 85 non-citizens who twice confirmed their status as non-citizens. A simple exercise with the binomial distribution shows that with 85 trials and a probability of success on each trial of 1.3233 percent, the probability of finding no successes is 32.2 percent. Thus, the observed outcome is one that is entirely plausible in the context of the frequency of non-citizen casting verified votes in 2010. It is an outcome that would occur nearly one third of the time if the actual non-citizen voting rate was precisely what the larger 2010 cross-sectional survey identified. Indeed, if one supposes that this estimate was a bit too high – that the true value was for instance the 0.65 percent Richman et. al. estimate would have been sufficient to account for Franken's 2008 MN win – one would get a null value when running 85 trials more than half (57.4 percent) of the time. A finding from a test with so little statistical power deserves at best only modest weight and attention as scholars assess the frequency with which non-citizens vote.[3]

2. **Flawed Measurement Error Assumptions**

This section shows that the assumptions that underlie the Ansolabehere et. al. (2015) argument do not hold when applied to the non-citizen voting case. As a result, their claim to have demonstrated that all or almost all apparent non-citizen voters are in fact citizens who erroneously claimed to be non-citizens does not hold up.

2.1. Hypotheses from the Measurement Error Assumption are Unsupported

If a finding based on analysis of a small subsample is purely the result of measurement error in group assignment, then there should be other observable implications that suggest auxiliary hypotheses to be tested. Tests of these hypotheses should lead to distinct conclusions depending upon whether measurement error is in fact responsible for a particular finding. In this case, those tests do not support the Ansolabehere et. al. argument.

The first place we turn is Table 2 of the Ansolabehere et. al. (2015) paper which directly contradicts the predictions that follow from their measurement error argument. The expectation that follows from their response error analysis is that nearly all of the individuals who listed their status as citizen in one year and non-citizen in another year are in fact citizens.[4] If that were the

---

[3] There is also some discussion of 2012 in the paper. The same analysis applies. Here they note that one non-citizen cast a validated vote, but did not state that he/she voted. The estimate of the voting rate of individuals who both said they voted and cast a validated vote in the Richman et. al. (2014) paper was 1.5 percent. Hence the probability of finding no such voters among 85 cases if the Richman et. al. estimate was precisely correct is 27.7 percent. Once again the critique lacks statistical power.

[4] Ansolabehere et. al. (2015, p. 409) argue that in any given year of the panel survey 19 citizens will (in expectation) erroneously state that they are non-citizens and one non-citizen will erroneously state that he or she is a citizen. Hence, of the individuals with inconsistent self-reported citizenship across the two waves of the survey roughly 38 of 40 should in fact be citizens. In fact the number of individuals who ever claim to be non-citizens in the 2010-2012 CCES panel is much lower than the 500 on which these extrapolations depend, so even less than the 5% estimated here should in fact be non-citizens.

case, we would expect to see these supposed citizens casting validated votes at a rate at least somewhat comparable to the rate at which other citizens cast votes. In fact, voting rates for individuals with inconsistent citizen-status self-reports in Table 2 of their paper are drastically lower than the voting rates of those who consistently identified as citizens (71.2% versus 7.1%). If their argument was correct, then the voting rate for these individuals would be approximately 68%. This should be a first warning that their claim concerning the frequency with which citizens erroneously identify as non-citizens is problematic.[5] Their argument is inconsistent with their data.

If Ansolabehere et. al. (2015) are right that all observed cases of non-citizens voting are the result of response error in the survey, this means that all individuals who were apparently non-citizen voters are citizens who erroneously claimed to be non-citizens. Likewise, this claim implies that all true non-citizens did not vote. This implies that seeming non-citizen voters should be similar to other citizen respondents. If Ansolabehere et. al. are correct, then when using a valid comparative metric (1) it should be possible to reject the null hypothesis that voting and non-voting non-citizens are the same (i.e. have statistically indistinguishable values of the comparative metric), and (2) it should *not* be possible to reject the null hypothesis that voting non-citizens and voting citizens are the same.

Arguably a valid set of questions for making this comparison can be found in the CCES question-battery asking respondents about attitudes toward immigration policy. Because they are personally impacted by immigration policy in a way that citizens are not, non-citizens should adopt distinctive immigration attitudes. Other survey datasets (e.g. Pew 2012) indicate that there are statistically significant differences in immigration attitudes between non-citizens and naturalized citizens and between non-citizens and all Latino citizens. If self-reported non-citizens who voted were in fact citizens who misstated their citizenship status, one would expect to see immigration policy responses in this subpopulation that strongly resemble those observed among citizens.

Arguably another valid indicator can be found in the CCES pre-election questions asking which presidential candidate the respondent preferred. In 2012 there were clear immigration policy differences between Obama and Romney which should have led immigrant non-citizens to be more likely to support Obama than other groups. Again, if Ansolabehere et. al. (2015) are correct that respondents who the survey indicated were immigrant non-citizen voters are in fact all citizens, we should see their responses resemble those of other groups.

---

[5] The analyses below use immigration attitudes to validate the identification of non-citizen status. There are no statistically significant differences (p<0.10) between the attitudes of consistent non-citizens (those who stated in both 2010 and 2012 that they were non-citizens, and the individuals who said they were a non-citizen in only one of the years. For five of six issues there are statistically significant differences between these inconsistent non-citizens and individuals who consistently stated that they were citizens. On average the attitudes of inconsistent non-citizens are sixteen points closer to those of consistent non-citizens than they are to those of citizens.

Table 1: Immigration Attitudes Among Self-Reported Citizens and Non-Citizens, 2012 CCES

(Numbers in parentheses are number of respondents in a particular category, e.g. total number of citizens in CCES.)

Question	All Citizens	Naturalized Citizens	Non-Citizens	Validated Non-Voting Non-Citizens	Validated Voting Non-Citizens	Degree to which non-citizens more pro-immigrant than citizens	Degree to which voting non-citizens more pro-immigrant than voting citizens	Degree to which non-citizens more pro-immigrant than naturalized citizens	Degree to which non-voting non-citizens more pro-immigrant than voting non-citizens
Grant legal status to all illegal immigrants who have held jobs and paid taxes for…	46% (53,622)	59% (2615)	68% (692)	65% (263)	69% (32)	22%*	23%*	9%*	-3%
Increase the number of border patrols on the US-Mexican border	57% (53,622)	45% (2615)	31% (692)	32% (263)	22% (32)	26%*	37%*	14%*	-10%
Allow police to question anyone they think may be in the country illegally	40% (53,622)	26% (2615)	19% (692)	21% (263)	25% (32)	21%*	17%*	7%*	4%
Fine US businesses that hire illegal immigrants	63% (53,622)	45% (2615)	34% (692)	38% (263)	34% (32)	29%*	32%*	10%*	-4%
Prohibit illegal immigrants from using emergency hospital care and public schools	32% (53,622)	21% (2615)	14% (692)	16% (263)	16% (32)	19%*	17%*	7%*	0%
Deny automatic citizenship to American-born children of illegal immigrants	37% (53,622)	24% (2615)	16% (692)	16% (263)	13% (32)	21%*	26%*	8%*	-3%
Percentage supporting Obama versus Romney two-candidate preferences only (pre-election survey)	54% (46,504)	68% (2,253)	80% (513)	76% (197)	92% (26)	26%*	41%*	11%*	-16%+

*Statistically significant difference p<0.001, + p<0.10, based upon chi-square test.

Table 1 compares the percentage responding yes to each question for five subsets of the sample: all self-reported citizens, naturalized citizens, all self-reported non-citizens, self-reported non-citizens who did not cast a validated vote, and self-reported non-citizens who cast a validated vote.

The analysis demonstrates that there are substantial and statistically significant differences (p<0.001 using a chi-square test) in the expected direction between self-reported non-citizens and citizens. In no case is this difference less than 19 percentage points. There are also substantial and statistically significant differences (p<0.001 using a chi-square test) between self-reported non-citizens and naturalized citizens, again in the expected direction. In no case is this difference less than seven points.

More to the point, if (as Ansolabehere and coauthors claim) all or nearly all voting non-citizens are citizens who misreported their citizenship status, then responses by non-citizens who voted would be quite different from those of other non-citizens – and these responses would be much more similar to responses by citizens. *The data in Table 1 are not consistent with this pattern.* In no case is there a statistically significant difference (p<0.05) between the immigration attitudes of non-citizens who cast a validated vote and non-citizens who did not cast such a vote. Indeed, in only one of the seven cases is even the direction of the relationship consistent with the hypothesized pattern. And the only instance with a difference on the margins of statistical significance (p=0.061) has a sign directly opposite of the one Ansolabehere, Luks, and Shaffner's argument would imply. By contrast, across all questions non-citizens who cast a validated vote had significantly more pro-immigrant attitudes than citizens.[6] The pattern of responses reported in Table 1 is inconsistent with the claim that self-reported non-citizens who cast validated votes were in fact citizens who mistakenly self-identified as non-citizens. Instead, this is the sort of pattern we would expect if these individuals were *all or almost all* actually the non-citizens they claimed to be.[7]

Other expectations that follow from the Ansolabehere, Luks, and Schaffner measurement error argument were examined by Richman et. al. (2014, pp 155-6) and received no support. Specifically, if their argument was correct then the racial demographics of non-citizen voters should resemble those of citizens. They do not. In addition, the geographic location of non-

---

[6] There are still several statistically significant differences if the analysis is repeated with a focus on the small group of non-citizens who both cast a validated vote and said they voted.

[7] These patterns are also inconsistent with the idea that self-reported non-citizen voters are individuals who are engaged in 'click through' without paying close attention to response categories. Click through ought to lead to a pattern of more random responses rather than responses that are systematically polarized. Furthermore, click through should generate lower levels of reliability in the immigration attitude scale among self-reported non-citizen voters. In fact the Crohnbach's Alpha coefficient for all self-reported non-citizens of 0.748 is virtually identical to the Alpha for non-citizen validated voters of 0.734 and the Alpha for non-citizen validated voters who also self-reported voting 0.743 and the Alpha for non-citizen self-reported voters of 0.785.

citizen voters should not be well predicted by the number of non-citizens living in different states. But it is.

2.2. Reasons for Response 'Error'

The Ansolabehere Luks and Shaffner analysis of inconsistent self-identification of citizenship status in the 2010 to 2012 CCES panel study assumes that the probability of a citizen misstating her status as non-citizen equals the probability of a non-citizen misstating her status as a citizen. If in fact non-citizens are much more likely to make 'errors' that misrepresent themselves as citizens than citizens are to erroneously claim to be non-citizens, then the inferences and arguments made by Ansolabehere et. al. (2015) are potentially no longer valid. We show here that decisions to obscure citizenship status likely account for a substantial portion of the supposed response error that forms the focus of the Ansolabehere et. al. analysis, thereby undermining their argument.

There are well known theoretical reasons to think that non-citizens are much more likely to misreport citizenship status than citizens are. Claiming to be a citizen (when not one) avoids any appearance of impropriety in contexts where revealing non-citizen status can be a legally sensitive issue.[8] By contrast it is difficult to think of circumstances in which an American citizen would have an incentive to lie about citizenship status while within the borders of the country. This means that non-citizens may be much more likely to waffle or masquerade when it comes to stating citizenship status in a variety of contexts. Demographic studies indicate that over-reporting of naturalization and citizenship by immigrants on surveys leads to significant discrepancies between naturalization records and census records (Van Hook and Bachmeier 2013). In the particular context of a survey about American politics the motive to misstate status is arguably greatest when other survey responses in conjunction with a citizenship-status statement in effect constitute an admission of vote fraud. Non-citizen voters have incentives to misrepresent either their citizenship status or their voting status. After all, claiming to be both a non-citizen and a voter is confessing to vote fraud, and the Federal Voter Registration Application specifically threatens non-citizens who register with a series of consequences. "If I have provided false information, I may be fined, imprisoned, or (if not a U.S. citizen) deported from or refused entry to the United States." This possible penalty would tend to reduce the proportion of non-citizens voters who would report having voted, and the portion of voters who would admit to being non-citizens.

Our core claim is that non-citizens are much more likely to make 'mistakes' when it comes to reporting citizenship status. A secondary claim is that such mistakes may be even more likely in contexts where admission of non-citizen status would constitute an admission of vote fraud.

---

[8] Indeed all non-citizens who register to vote have lied about citizenship status as federal and state forms require that individuals attest to their citizenship. Substantial numbers of non-citizens have been identified on voter registration rolls. For example see Richman 2016a and 2016b.

If in fact non-citizens are much more likely to accidentally and/or intentionally claim to be citizens than citizens are to accidentally claim to be non-citizens, this should be apparent across repeated measures in the 2010 through 2014 CCES panel. The relevant quantities are conditional probabilities – the probability that a respondent, having stated a particular status in two of the three panels, will state a different status in a third panel. We expect to observe a much higher rate of stating a different status for those who twice stated they were non-citizens than for those who twice stated they were citizens.[9]

The strongest comparisons are those involving individuals who reported that they were citizens in 2010 and 2012 and individuals who reported they were non-citizens in 2012 and 2014.[10] In both cases there is no commonly experienced change in legal immigration or citizenship status that could account for survey response error in the third year.[11] Hence, almost any deviation from consistency in the third year (2010 for twice-asserted non-citizens and 2014 for twice-asserted citizens) can only be accounted for on the basis of unintentional or intentional measurement error.

Table 2: Three Wave Citizenship Status Response Consistency in the CCES

	Citizen in 2014	Non-Citizen in 2014	Portion inconsistent in third measurement
Claimed to be a citizen in 2010 and in 2012	9426	4	0.00042
	Citizen in 2010	Non-Citizen in 2010	
Claimed to be a non-citizen in 2012 and 2014	3	23	0.13

---

[9] Before proceeding further, we pause to note that the argument made by Ansolabehere et. al. (2015) is based on an analysis of a substantially smaller dataset than the original Richman et. al. 2014 study. Ansolabehere et. al. (2015) base their response error measurements on a comparison of citizenship-status self-reports in the 2010 and 2012 waves of the CCES panel study. Their critical results involve 56 respondents who gave inconsistent responses – claiming to be citizens in one year of the study and non-citizens in another year, and 85 respondents who consistently stated that they were non-citizens. These are relatively small numbers. Hence, readers should prepare themselves for further analysis of small subsamples of sub-samples, as we will need to reanalyze these and similarly small groups as we point to the flaws in the conclusions drawn. If their critique, based as it is on such small samples, has any validity, then our response much join it on this terrain.

[10] A similar pattern emerges in the other possible comparisons as well.

[11] Renunciation of US citizenship could theoretically account for some of the observed error among twice-reported citizens. If present, this would lead to an even higher difference in group reliability estimates – it would lend further support to our position.

Table 2 reports three-wave response consistency in the 2010 through 2014 CCES panel study. As expected, citizens have a much higher reliability than non-citizens. For individuals who stated they were citizens in 2010 and 2012, a consistent response was provided 99.958 percent of the time in 2014. By contrast, for individuals who twice stated they were non-citizens in 2012 and 2014, a consistent response in 2010 was provided only 86.96 percent of the time. The difference between these proportions is statistically significant with a difference of proportions z-test (p<0.05). This rather strongly suggests that the reliability estimate by Ansolabehere et. al. (2015) was biased downward by the much lower reliability of self-reported citizenship status among non-citizens.

Our second expectation involves a pattern of claiming citizenship status when voting. If inconsistent reporting of citizenship status reflects in part lying about citizenship to avoid the appearance of illegal activity, then we would expect the following pattern: among individuals who once reported they were citizens, and once reported they were non-citizens, the probability of casting a vote should be higher in the year when they reported they were citizens. Although the sample sizes are small and the differences do not all reach standard levels of statistical significance, there is some evidence of this pattern in the data. In Table 2 of Ansolabehere et. al. 15 percent of inconsistent respondents who claimed to be citizens in 2010 cast validated votes whereas only 2.8 percent who claimed to be non-citizens in 2010 cast validated votes (p = 0.12 two-tailed Fisher's Exact Test). Self-reported voting follows a similar pattern. 50 percent of respondents who claimed to be citizens in 2010 and then non-citizens in 2012 reported voting in 2010 compared to only 25 percent reported 2010 turnout among those who in 2010 claimed to be non-citizens and in 2012 claimed to be citizens (p = 0.08 two-tailed, Fisher's Exact Test). There were 14 individuals who said they were non-citizens in 2012 and voters in 2010. In 2010 71 percent stated they were citizens. As modest evidence of misstating to avoid admitting vote fraud, we note that in 2012 when these individuals said they were non-citizens their reported electoral participation rate dropped by 43%, a statistically significant decline (p=.016 Fisher's exact test).

The key implication is that a large portion of the respondents with inconsistent citizenship self-reported status are in fact likely to be non-citizens. It follows that the expected portion of respondents in the CCES cross-sectional surveys who are citizens and misreport that citizenship status as non-citizen is substantially lower than the estimates reported by Ansolabehere et. al. (2015) imply. This directly undermines their inferences concerning whether citizens who erroneously claim to be non-citizens are sufficiently numerous to account for observed levels of voting by self-reported non-citizens in the CCES, as will be shown next.

2.3. Consequences of Revising the Reliability Estimate

The revised estimate of the frequency with which citizens misidentify as non-citizens makes a significant difference for the inferences one draws from cross-sectional CCES data of the sort examined by Richman et. al. (2014). Consider for instance the 2012 CCES cross-sectional survey. In the 2012 CCES cross-sectional survey 32 respondents who identified as non-citizens cast a verified vote. If we assume that the portion of citizens erroneously reporting that they are non-citizens is that estimated in the first row of Table 2, then we are in a position to

estimate the probability that 32 citizens with verified votes erroneously misstated their citizenship to account for the entirety of the apparent electoral participation by non-citizens. This is the claim made by Ansolabehere et. al. (2015) and the revised reliability estimates undermine it.

Table 3. Estimated Voter Turnout by Non-Citizens in 2012 CCES Cross-Section (Number of voters/total in sample in parentheses.)	
Self-reported voting as a percentage of all non-citizens	8.8% (61 of 692)**
Validated voting as a percentage of Catalist matched respondents	10.8% (32 of 295)*

** Binomial probability result generated entirely by citizen response error <0.000001.
* Binomial probability result generated entirely by citizen response error <0.0005.

Table 3 reports the number of self-reported non-citizens who cast validated votes and self-reported votes, and the probability that these estimated levels of non-citizen voting could be accounted for entirely by response error on the part of citizens. The math is straightforward. For instance, 81 percent of self-reported citizens with a Catalist-file match voted in 2012. Thus, the probability that any given citizen will both have a verified vote and have erroneously stated non-citizen status is only 0.00034. Working out the binomial probabilities across all 45221 respondents with a voter file match yields a probability of only 0.00017 that 32 or more such individuals were present in the 2012 survey. Hence, by our estimate the probability is very small indeed that all of the instances of self-reported non-citizens who cast verified votes in the 2012 cross-sectional CCES survey were in fact instances of citizens who cast a verified vote and misstated their citizenship status.

The conclusion by Ansolabehere et. al. (2015) that "the likely percent of non-citizen voters in recent US elections is 0" depends upon what was then an untested estimate of the reliability of citizenship status self-reports by citizens. With a corrected measure of citizenship status self-report reliability made possible by the 2010 through 2014 CCES panel study, measurement error in group assignment cannot account for the level of participation among self-reported non-citizens observed in the CCES cross sectional survey.

3. A More Complete Set of Inferences

This section stands independent of the analysis we offered in the previous section. Here we assume that Ansolabehere et. al. (2015) are entirely correct about the frequency with which citizens erroneously claim to be non-citizens. Our aim in this section is to show that even if one accepts their argument, their conclusions are incorrect – the CCES provides significant evidence of non-citizen involvement in the U.S. electoral system.

3.1. Correlated Higher-Frequency Events

As discussed above, Ansolabehere et. al. estimate the reliability of the citizenship status measure, and conclude that citizens would make enough errors on the citizen-status question to account for the observed level of validated voting by self-reported non-citizens in the CCES.

However, their error estimate is too low to account for the observed rate of voter registration among non-citizens in the CCES. This strongly suggests that non-citizens do register to vote in US elections.

Our approach is to analyze higher frequency behaviors that correlate with the behavior or interest. To the extent that such behaviors occur at a rate too high to be accounted for by group assignment measurement error, they provide another way to infer the presence of particular activities. We consider voter registration as a candidate measure. In all US states save North Dakota, registration is a precondition for electoral participation. Hence, registration to vote necessarily occurs at a higher frequency than voting.

Table 4. Estimated Registration by Non-Citizens
(Number of individuals registered divided by sample size in parentheses.)

	(1) 2012 Cross-Section	(2) 2012 Panel (test-retest non-citizens)	(3) 2014 Panel (test-retest-retest non-citizens)
Self-reported registration as a percentage of all non-citizens.	14.5% (100/692)**	14.2% (12/85)**	13.0% (3/23)**
Validated registration as a percentage of Catalist matched respondents.	22.0% (65/295)*	10.6% (5/47)**	6.3% (1/16)**

** Binomial probability that this result could have been generated entirely by citizen response error <0.000001.
* Binomial probability result generated entirely by citizen response error <0.05.

Table 4 reports analysis of the frequency of voter registration (self-reported or Catalist verified) for the 2012 cross-sectional as well as the 2012 and 2014 panel studies. As discussed more thoroughly below, although the sample size in the panel study is smaller, it offers the advantage that we can be very confident that individuals are in fact non-citizens as they twice (2012 panel) or thrice (2014 panel) repeated that they were non-citizens.

Estimates of binomial probability that the observed results reflect citizenship self-assignment error use the reliability estimate calculated by Ansolabehere et. al. (2015) instead of the corrected measure we suggest in the previous section. Ansolabehere et. al. report that the citizenship status question on the CCES has a reliability of 99.9 percent.[12] If 99.9 percent of responses to this question are reliable, this suggests that the chances of an error being made twice – in particular a citizen responding twice that he or she was a non-citizen – is $(1-.999)^2 = 0.000001$. In the large set of survey respondents to the CCES, we can use the binomial distribution to model this process of a citizen randomly making (or not making) twice a mistaken response to the citizenship question. The cumulative binomial distribution can be used to

---

[12] Although we present evidence above that this estimate was likely too low for citizens and too high for non-citizens, this section works on the basis of their original measurement.

calculate the probability that a particular outcome or set of outcomes will occur. In particular, our interest is in the probability that a particular number of citizens will repeatedly make the mistake of asserting that they are non-citizens. To take an example, consider that in the 2010-2012 panel there are 18,878 respondents who each either made this mistake twice or not. The binomial probability that no citizen will twice misstate his or her citizenship status is very high even across 18,878 trials (98.1 percent), and the probability of at least one respondent who twice indicated he or she was a non-citizen in fact being a citizen is low: 0.0189. The likelihood is therefore very high that all of the respondents who twice indicated they were non-citizens in the 2010 to 2012 CCES Panel (Column 2 of Table 4) were in fact non-citizens. And the probability is even higher that all of the respondents who three times reaffirmed that they were non-citizens (Column 3 of Table 4) were in fact non-citizens.[13]

In each column the pattern is consistent – more registration is observed than can be accounted for by the Ansolabehere et. al. (2015) estimate of the reliability of citizen status self-reporting.[14] Thus, the evidence of response bias in citizen-status self-assignment cannot account for the observed level of voter registration among non-citizens. Since registration is a precondition for and correlate with voting, this provides indirect evidence that non-citizens participate in U.S. elections.

One potential rejoinder would be to note the possibility that Catalist mismatched all of the non-citizens with validated registration status. This possibility is particularly remote when those individuals also reported that they were registered. For 2012, two of the test-retest non-citizens with validated registration status also self-reported that they were registered to vote and in 2014 the test-retest-retest non-citizen with validated voter status also indicated that he or she was registered. Note that this is an individual with a very high probability of being a non-citizen as non-citizen status was reconfirmed in 2010, 2012, and 2014. As noted in the table the probability that this individual was a citizen who thrice randomly misstated citizenship status is, on the basis of the Ansolabehere et. al. (2015) reliability estimate, less than 0.000001. Obviously these are very small numbers, but they help make the point nonetheless, as a single valid case is sufficient to prove existence. For these individuals we can be even more confident that they were in fact genuine non-citizen registrants.

3.2. Test-Retest Reliability in Voting

We have already begun to introduce the final strategy for addressing the risk of group assignment bias – to focus on respondents for whom repeated measurement of group membership allows for more confident group assignment, as we applied it to voter registration in the preceding paragraphs. As should already be clear from the discussion above, participation by

---

[13] A potential objection might be that the likelihood of a second error following the first is higher because of individual idiosyncrasies that made the first error more probable. This objection potentially weighs against the test-retest and test-retest-retest analysis, but it has no bearing when it comes to the analysis of the 2012 cross-sectional survey.

[14] Obviously if the adjusted reliability estimate for citizens proposed in the section below was used instead, these results would be even more strongly statistically significant.

even a few test-retest non-citizens in the CCES sample presents a major problem for the claim by Ansolabehere et. al. (2015) that no non-citizens participate in US elections.

Table 5. Estimated Voter Turnout by Non-Citizens (Number of voters / total sample parentheses.)	2012 Panel (test-retest)	2014 Panel (test-retest-retest)
Self-reported voting as a percentage of all non-citizens	11.8% (10/85)**	8.7% (2/23)**
Validated voting as a percentage of Catalist matched respondents	2.1% (1/47)*	0% (0/16)

**Binomial probability result generated entirely by citizen response error <0.000001.
*Binomial probability result generated entirely by citizen response error <0.05.

Ansolabehere et. al. (2015) do consider participation by such test-retest non-citizens. Table 2 of their paper focuses on validated voting in the 2010 election. This is convenient for their argument, as none of the four non-citizens with validated voter-registration status in 2010 cast a validated vote. A display of the same table for 2012 would have provided less support for their claim. In the 2012 election, one of the five test-retest non-citizens with validated voter registration status cast a validated vote. Table 3 of this paper provides this data. The probability that this validated vote was cast by a citizen rather than a non-citizen is quite low. Even with 17,831 respondents with a Catalist match, the cumulative binomial distribution gives probability of one or more false positives arising from measurement error on the citizenship question as only 0.015.[15]

Table 5 also examines self-reported voting among test-retest non-citizens. Among the 85 test-retest non-citizens in the 2010-2012 CCES panel, all were asked if they voted in 2010, and 15 were asked if they voted in 2012. In 2010 six (7.1 percent) selected the "yes I definitely voted" option, in 2012 ten (11.8 percent of the 85) selected the "I definitely voted" option, and in 2014 two of the 23 individuals (8.7 percent) who had thrice indicated they were non-citizens selected the "I definitely voted" option. In all cases the probability that these results merely reflect response error on the immigration status question by citizens is vanishingly small (p<0.000001), even using Ansolabehere et. al.'s arguably biased measure of the reliability of citizen status self-reports. Some individuals who are most likely non-citizens clearly do report that they are voting in U.S. elections.

We note in passing that other survey responses sometimes provide opportunities to re-measure citizenship status in the 2012 cross-sectional study. For example, when asked why they

---

[15] 87.1 percent of respondents in the overall survey who had a Catalist match cast a verified vote. Therefore the probability of any given survey respondent being a citizen who twice reported being a non-citizen and cast a verified vote is only 0.000000871.

did not self-report voting, a substantial number of self-identified non-citizens indicated that the reason was that they were "not a citizen" or some variant thereof. Open ended questions in the 2012 CCES invited respondents who indicated some "other" reason for not voting to provide up to two explanations for the decision to not vote. A substantial number of self-reported non-citizens indicated that they had not voted because of their immigration status (i.e. "not a citizen" or "no soy ciudadano," "have a green card" or "permanent resident", or "I do not have my GC yet"). Of the 412 self-reported non-citizen respondents asked why they did not vote almost half (47%) indicated that their non-citizen status was a reason for not having voted. A high level of confidence is warranted that these 192 respondents are indeed non-citizens as they at least twice indicated their citizenship status, including at least once in an open ended response. Catalist found a file match for 102 of these repeatedly self-identified non-citizens. And despite it being nearly certain that they were in fact non-citizens, 11 (10.8%) had active voter registration status, and two of the 102 (1.96%) cast validated votes. One of these respondents was explicit that although registered there was no intention to cast a vote. "I am not a U.S. citizen, but was mistakenly sent a voter registration card anyway. Will not take advantage of mistake to vote illegally." We see no way to dismiss evidence such as this of non-citizen registration.

4. Conclusion

As Richman et. al. (2014) had noted and addressed in the original article, Ansolabehere et. al. (2015) make a useful general point – that group-membership measurement error rates must be considered very carefully when analyzing small subsamples. To that end Richman et. al. (2014) had examined racial demographics, geographic location, and immigration attitudes of non-citizens who self-reported voting. This paper takes that validation effort several steps farther.

Section 2 brought to bear multiple lines of evidence that Ansolabehere et. al. (2015) are simply wrong about the frequency with which citizens erroneously claim to be non-citizens. Self-reported non-citizens have attitudes toward immigration that are entirely inconsistent with the claims made by Ansolabehere, Luks, and Schaffner that they are citizens. And there are good reasons to believe that the significantly higher error rate by non-citizens on the citizen status question undermines the argument of Ansolabehere et. al. . Our rejoinder provides reason to believe that a substantial number of the self-reported non-citizens who voted were in fact non-citizens.

Setting aside our evidence from Section 2, Section 3 assumed that Ansolabehere et. al. are correct about the rate of response error in the citizenship status self-report question in the CCES. The first sub-section examines the voter registration data that Ansolabehere et. al. ignored in their critique. We show that the voter registration data is flatly inconsistent with their claim that zero non-citizens participate in US elections. We also re-examined responses by test-retest non-citizens, and find significant evidence contradicting the claim made by Ansolabehere and colleagues that none vote in U.S. elections.

We have shown that each of four independent approaches to evaluating electoral participation by non-citizens indicates that in fact a small number of non-citizens do most likely

participate in US elections. Analysis of group-specific error rates; repeatedly measured individuals; higher frequency behaviors; and hypotheses that follow from the assumption that responses are driven by group-identification errors all yield the same independent conclusion: a refutation of the Ansolabehere et.al. (2015) contention that the Richman et. al. (2014) non-citizen participation results "are completely accounted for by very low frequency measurement error" among citizens. Their assertion that they have "debunked" that paper has no basis in the data.

Although the criticisms of our work speak to the inherent difficulty of studying individuals who face strong pressures to misrepresent their interests and behaviors, we stand by the basic claims of Richman et. al. 2014. A more thorough analysis of the data makes clear that response error in the citizen-status question cannot account for the observed level of non-citizen verified and reported voting in the CCES. Hence, the CCES survey does provide substantial evidence that in the United States non-citizens hold verified registration status, cast verified votes, report they are registered, and report they are voters.

Works Cited

Ansolabehere, Stephen, Samantha Luks, and Brian F. Schaffner. 2015. "The perils of cherry picking low frequency events in large sample surveys." *Electoral Studies*. 40, 409-410.

Richman, Jesse T., Gulshan A. Chattha, and David C. Earnest. 2014. "Do non-citizens vote in US elections? *Electoral Studies.* 39, 149-57.

Richman, Jesse. T., 2016a. DACA Non-Citizen Registration Rate Estimate for North Carolina. Downloaded on 12/8/2016 from https://fs.wp.odu.edu/jrichman/2016/10/20/daca-registration-rate-estimate/

Richman, Jesse T. 2016b. Non-Citizen Terminated Registration Rates in Virginia Counties. Downloaded on 12/8/2016 from https://fs.wp.odu.edu/jrichman/2016/11/05/non-citizen-terminated-registration-rates-in-virginia-counties/

McLaughlin, John. 2013. "National Hispanic Survey Results" Presented June 21st 2013. Downloaded February 1, 2017 from http://www.mclaughlinonline.com/lib/sitefiles/National_Hispanic_Presentation_06-21-13_-_FOR_RELEASE.pdf

Pew. 2012. Bilingual dual-frame (cell phone and landline) telephone survey of Latino adults residing in the U.S., conducted September 7, 2012-October 4, 2012.

Schaffner, Brian. 2016. "Trump's Claims About Illegal Votes Are Nonsense. I Debunked the Study He Cites as 'Evidence.' The real number of non-citizen voters is more like zero." Downloaded on December 8, 2016 from http://www.politico.com/magazine/story/2016/11/donald-trump-illegal-votes-evidence-debunked-214487

Van Hook, Jennifer, and James D. Bachmeier. 2013. "How well does the American Community Survey count naturalized citizens?" *Demographic Research*. 29(1) pp. 1-32. DOI: 10.4054/DemRes.2013.29.1