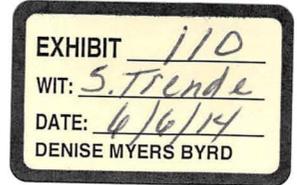EXHIBIT F

Sweeping Conclusions From Census Data Are a Mistake

By **Sean Trende** - May 9, 2013

The big news of the week in the elections world has nothing to do with Jodi Arias or Benghazi. It has to do with the latest release from the U.S. Census Bureau about the 2012 electorate: the Current Population Survey data, or CPS for short. This is a November survey that goes out with the one used to measure the unemployment rate and other statistics. This survey, however, asks people if they voted. Its main benefit is that it has a large sample size, which allows for particularly fine-grained analysis.

The recent release has been reported by others with banner headlines like this one: "For First Time on Record, Black Voting Rate Outpaced Rate for Whites in 2012." A flood of analysis has predictably followed, ranging from deep dives into the data, to commentary on the problems it poses for Republicans, to commentary on the problems it poses for Democrats if this level of African-American enthusiasm proves unsustainable after Obama.

Here's my take: Analysts should be much more circumspect in interpreting these data. The report is clearly off in an important respect. Depending on the cause of the error, it could flow through and affect most of the other findings of the report.

The problem is pretty straightforward. On Table 2, which you can read here, the CPS data conclude that there were 1.4 million more Hispanics who voted in 2012 than in 2008, 547,000 more Asians, 1.7 million more blacks, and 2 million fewer whites. That works out to a total of 1.8 million more votes cast in 2012 than 2008, according to the CPS survey.

But if there is one thing that we absolutely know about 2012, beyond any reasonable doubt, it is that turnout actually dropped from 2008. In fact, it dropped substantially. Dave Wasserman followed the 2012 returns as closely as anyone, and

he calculates that turnout dropped from 131,313,820 in 2008, to 129,069,194.

So, the CPS data say that there were around 4 million more votes cast in 2012 than was actually the case. This means that those voting numbers we talked about two paragraphs earlier actually have to be reduced, in some combination, by a total of 4 million.

How should they be reduced? That's the bigger problem. Perhaps the easiest method is to assume that the census data overstate each population group proportionally. In other words, of the 4 million extra votes, 73 percent should be ascribed to whites, 13 percent to blacks, and so forth. If we use this method of dealing with the overcount, then there were actually 5 million fewer white voters than there were in 2008, 1.1 million more blacks, 211,000 more Asians, and 1.3 million more Hispanics. That is a significant difference.

But how much confidence should we have that all groups overstated their voting patterns equally? That strikes me as more likely than the possibility that one group in particular overstated its voting participation, and that none of the other groups did so, but it's impossible to quantify how much more likely the first possibility is.

If it is instead true that whites, blacks, Asians and Latinos overstated their participation at different rates, then the whole exercise becomes very difficult to sort through. Perhaps 80 percent of the 4 million vote overcount should be ascribed to white voters. Perhaps 40 percent of the overcount should be. We just don't know, nor can we even state with much confidence how likely it is that a given apportionment of the overcount is correct. All we know is that there are millions of possible combinations to choose from.

If it is the case that racial groups overstated their participation at different rates, then it presents a problem that pervades all of the data in the report. Let's say, for example, that whites disproportionately overstated their actual voting rate in 2012. That would also mean that the age cohorts are skewed, because whites are disproportionately older.

It would also affect the headline data. Let's say instead that whites disproportionately understated their actual voting rate in 2012, and all other groups overstated it. This could mean that African-Americans did not, in fact, vote at a higher rate than whites this year.

In fact, we can't even be certain that the problems are with the current year. It might be that African-Americans disproportionately understated their performance in 2008 but did not in 2012, which would make it seem as though there was a disproportionate increase in votes reported this time around.

One intriguing possibility along these lines, suggested by the Huffington Post's Michael McDonald, is that non-response bias accounts for the discrepancy. The census treats someone who doesn't respond as not having voted. In 2012, blacks increased the rate of response more than any other group, which might have created a disproportionate increase in reported black turnout.

Rather than treating non-respondents as "did not vote," McDonald simply drops them from the data sets. He then does something very illuminating: He re-creates the age cohorts with his assumptions about non-respondents.

The differences are aren't huge, but they are important: He finds that 18- to 24-year-olds dropped off by 8.4 percent, as opposed to the 7.3 percent growth that the census reported, and that 75-year-olds increased their participation rate by 0.9 percent, rather than 2.2 percent. He finds that the decrease in participation was, in fact, disproportionately concentrated among whites (both Hispanic and non-Hispanic). Perhaps most interestingly, by re-creating the 2008 data set, he concludes that African-American participation outpaced white participation in 2008 as well.

This is an intriguing theory, but the assumption that we can drop non-participants without affecting the data isn't clearly correct, although it is clearly reasonable. But it might also be the case that people who don't respond to the survey are actually substantially more likely not to have voted than those who do respond; dropping them would skew this effect.

Before closing, I want to emphasize that it's not clear that the census did anything wrong. Its results are not "garbage-

in/garbage-out." The over-response issue is pervasive, and it is something of which researchers are keenly aware. This particular discrepancy has even popped up before: The census data found an increase of 5 million votes between 2004 and 2008, when the actual increase was 9 million. It wasn't as salient, because at least the actual results and the exit polls both saw increases in vote totals, but it was there. (There are also good years; the increase from 2000 to 2004 was pretty close.) Perhaps because of this, the census report comes with a much less sexy headline than has been reported: "The Diversifying Electorate -- Voting Rates by Race and Hispanic Origin in 2012 (and Other Recent Elections)."

But the bottom line is this: We know that there's a problem with the data here, at least in terms of how people are reporting it out. The bigger problem is that we don't know exactly what that problem is. The data aren't useless by any stretch, and exit polling has its own, probably greater, problems. But because of this known issue, analysts and reporters should avoid making sweeping pronouncements on the basis of these data. There's just too much that we don't know.

//

Sean Trende is Senior Elections Analyst for RealClearPolitics. He is a co-author of the 2014 Almanac of American Politics and author of The Lost Majority. He can be reached at strende@realclearpolitics.com. Follow him on Twitter @SeanTrende.

Page Printed from:
http://www.realclearpolitics.com/articles/2013/05/09/sweeping_conclusions_from_census_data_are_a_mistake.html
at May 12, 2014 - 04:02:06 PM CDT