JUST FACTS.
SEIZE THE DATA

Supplemental Documentation

Article: "Critics Fail to Debunk Explosive Study on Illegal Voting by Non-Citizens"

By James D. Agresti

Published June 27, 2024 at https://www.justfactsdaily.com/critics-fail-to-debunk-explosive-study-on-illegal-voting-by-non-citizens

Just Facts Previously Notified Snopes of Its Errors Back in 2017

7/11/2017 9:41 AM
http://www.snopes.com/contact/

Correction Request: "Did 5.7 Million 'Illegal Immigrants' Vote in the 2008 U.S. Election?"

Dear Snopes:

Our institute has found materially misleading assertions in your post "Did 5.7 Million 'Illegal Immigrants' Vote in the 2008 U.S. Election?." We have published an article that addresses these falsehoods, and I am writing to make you aware in case you would like to respond or issue a correction: http://www.justfactsdaily.com/false-arguments-against-evidence-of-vote-fraud/

Sincerely,

James D. Agresti | President | Just Facts | 641 Shunpike Rd #286 | Chatham, NJ 07928 | www.justfacts.com
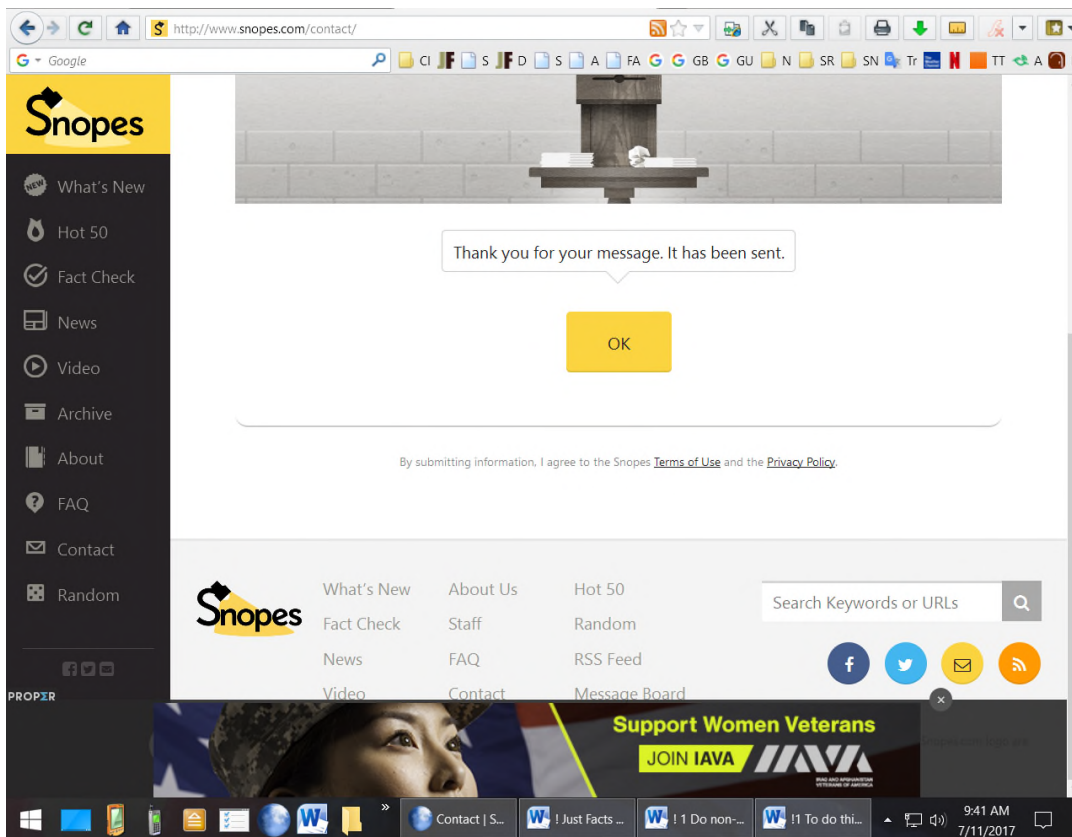
Email From Kasprak to Agresti on 5/16/24

On Thursday, May 16, 2024 at 04:07:51 PM CDT, Alex Kasprak <XXX@snopes.com> wrote:

Hello Mr. Agresti,

As you know, I am a writer for the fact checking website Snopes. We are writing a short piece on your May 12 Just Facts report asserting, based on your "enhanced analysis" of Richman et al (2014), that "10% to 27% of non-citizens are illegally registered to vote."

My criticism is no different now than it was in 2017, when we wrote a piece about your use of that paper to support the claim that "as many as 5.7 million illegal immigrants might have voted in the 2008 election."

Any use of that paper, even with the additional data produced since then, stems from a pool of at most 71 purported non-citizen respondents, and extrapolations based on these numbers, known to be at least in part generated by incorrect responses, to the entire U.S. population is wildly untenable.

As your colleague Brian Schaffner wrote to us in 2017:

> I don't know any serious survey researchers who would have tried to extrapolate 100 or so respondents from a large survey like this to produce a range that large without tracking back to think about the dubiousness of that projection.

> I am wondering if you have any response to that statement in general, or to your critics' points about the high rate of potential survey response error, at least in the Harvard/YouGov polls.

We are seeking a response by 5:00 pm Pacific time today.

Best,

Alex Kasprak

Email From Agresti to Kasprak on 5/16/24

5/16/2024 10:08 PM

Hi Alex,

Feel free to call me Jim.

I just found your email in my spam filter. How much time was there between when you sent this email and the 5:00 PST response time that you requested? Surely you must know that people with serious responsibilities can rarely respond on such short notice.

I wrote a [thorough rebuttal to your 2017 piece](#). Were you unaware of it? It is the first link in the following paragraph of my [new study](#):

> So-called fact checkers and certain scholars have repeatedly tried to dispute the *Electoral Studies* paper and Just Facts' study. However, their criticisms were [mathematically illiterate](#) and laced with [unrealistic assumptions](#), [empty arguments](#), [half-truths](#), and [outright falsehoods](#).

My 2017 rebuttal is still applicable, especially these facts:

> Both Snopes and PolitiFact are making a simplistic math error, for as the textbook *[Statistics for K–8 Educators](#)* explains:

> > It is remarkable that the margin of error when you estimate a percentage depends only on *p* [the proportion of people who answer a poll question in a certain way] and *n* [the number of respondents]. This explains why a national random sample of 1,000 people can accurately represent 200 million people. The *p* from the sample of 1,000 people is not likely to be more than 3% off from what you would find if you did include all 200 million.

> …

> To be specific, for a sample of 339 people, the difference in the [margin of sampling error](#) (with at least 95% confidence) for a population of 2,000 people versus 20 million people is less than a single percentage point (±4.8 versus ±5.3). …

> PolitiFact and Snopes also misconstrue the implications of the number of voters in the survey. For a given sample size and margin of error, confidence in the results actually increases when the portion of people who answer a poll question in a certain way is well below or above 50%, as is the case with this poll data. Per the textbook *[Mind on Statistics](#)*, in such instances the "conservative approximation of the margin of error … usually overestimates the actual size of the 95% margin of error, and thus leads to an underestimate of how confident we can be that our interval estimate covers the truth."

In this sample of 339 non-citizens, the ±5.3 sampling error shown above would have 95% confidence if the data indicated that 170 people voted, but since it indicates that 38 people voted, it has 99.8% confidence. Yet, Snopes and PolitiFact mislead their readers to believe that 38 voters in the survey makes the data pointless. …

Also, Snopes, PolitiFact, and the *Huffington Post* repeatedly warn that Just Facts' study is based on an "extrapolation." In reality, the study is based on a straightforward application of survey data—and these same organizations along with every major media outlet routinely cite similar figures without calling them "extrapolations."

Because it is often impractical to collect information from every person in the United States, governments frequently obtain key data through surveys. This is true of official government data on crime, education, employment, the economy, and an enormous array of Census data. These surveys typically use much larger samples than the one used by Just Facts, and thus, the margins of sampling error are smaller, but the basic method and principle are the same: survey a representative sample and apply the results to the U.S. population.

An extrapolation, on the other hand, is estimating "the value (of a function or quantity) outside a range in which some values are known" (*American Heritage Dictionary of Science*, 1986). This conveys a form of uncertainty that is not present in Just Facts' study. One could broadly parse this definition to include Just Facts' study, but then it would also apply to reams of government data that are rarely (if ever) called extrapolations.

In addition to the above, the claim that the previous data was "known to be at least in part generated by incorrect responses" was based on flawed assumptions that I documented in 2016. Moreover, it has since been totally discredited by pages 73–74 of Richman's 2023 report, which I linked to in my new study in the following sentence:

"Enhancements over previous studies include … multiple citizenship questions in the survey limit the possibility of honest mistakes by survey respondents."

If you honestly want to check the facts, read those two pages at the link.

Best,

Jim
James D. Agresti | President | Just Facts | 3600 FM 1488 Rd. | Suite 120 #248 | Conroe, TX 77384 | justfacts.com

Email From Kasprak to Agresti on 5/20/24

On Monday, May 20, 2024 at 04:32:05 PM CDT, Alex Kasprak <XXX@snopes.com> wrote:

Hello Jim,

Thanks for your response and I apologize for the deadline I gave you previously. At the time I was unaware how much time I would be given to pursue the story, but either way it was not sufficient for a serious response.

The primary argument in your 2017 rebuttal piece seems to be that extrapolation is appropriate based on the statement in a textbook that reads in part "a national random sample of 1,000 people can accurately represent 200 million people." I have read your argument against the sampling in the broader CES study, but what leads you to believe you have anything close to approaching a national random sample of the non-citizen population underlying any of your extrapolations?

In corresponding with Dr. Richman, he had this to say of your 10-27% range of noncitizen voter registration:

> I think the 10 percent figure has a foundation in my work, though it needs to be caveated more clearly to indicate the risks of being biased upwards, and I think it is absolutely inappropriate as a lower bound.

> The appropriate lower bound should be based upon individuals with validated registration status who said they were registered (i.e. less than one percent in 2022). I don't think the 27 percent figure is reasonable, partly because of changes in the methodology of the CES, and partly because I don't fully understand how it was calculated. [...]

> In my view, the adjustment strategy we applied [in 2014] was more appropriate in the earlier years of the data when Catalist was matching CES respondents not only with the voter file, but also with a variety of other commercial databases. I think it makes little sense in the context of the Target Smart voter file match which only identifies individuals who are registered.

> Thus, I'm unsure how the 27 percent number came about, I don't think it has a foundation in my methodology, and I don't think it's plausible in light of various other indicators of the frequency of non-citizen registration and voting.

I wanted to give you an opportunity to respond to his comments as well.

I'm looking to submit by the end of my day tomorrow. If that is unreasonable timing, however, please let me know.

Best,

Alex

Email From Agresti to Kasprak on 5/21/24

5/21/2024 2:47 PM

Alex,

Thanks for your reply and apology.

The answers to your questions and Richman's comments were already documented in my earlier email and publications. To reiterate:

- Describing Just Facts' studies as an "extrapolation" is inapt and misleading.

- The data used for the 2014 *Electoral Studies* paper and the 2017 Just Facts study were both weighted to make the results nationally representative of the *non-citizen* population.

- The data used for the 2023 Richman study and the 2024 Just Facts study were matched and weighted to make the results nationally representative of the *citizen* population, which may *understate* the non-citizen registration rate.

- Richman's "appropriate lower bound" is based on an unjustified counterfactual methodology that excludes all non-citizens who:
  - *were* matched to a voter registration database but stated in the survey that they *aren't* registered to vote.
  - *weren't* matched to a voter registration database but stated in the survey that they *are* registered to vote.

- Richman's claim that the 10% lower bound from Just Facts' 2024 study "needs to be caveated more clearly to indicate the risks of being biased upwards" is belied by the following facts:
  - Just Facts thoroughly caveated all aspects of the study.
  - The caveats suggest that the 10% figure is more likely to be biased downwards instead of upwards.

- Richman's claim that the 27% upper bound from Just Facts' 2024 study isn't "plausible in light of various other indicators of the frequency of non-citizen registration and voting" is belied by these facts:
  - His own 2014 *Electoral Studies* paper found that "roughly one quarter of non-citizens" in the U.S. "were likely registered to vote."
  - The "other indicators" to which he refers were based on biased methodologies that systematically lowball registration rates. Moreover, Richman fails to reveal these caveats when calculating these rates.

- Richman's claim that the 27% upper bound from Just Facts' 2024 study doesn't have "a foundation in my methodology" is belied by these facts:

- The study's spreadsheet proves that it does have a foundation in the methodology used by Richman and his coauthors in their 2014 *Electoral Studies* paper.
- Richman used and defended the same methodology in 2016 and 2017.

Best,

Jim
James D. Agresti | President | Just Facts | 3600 FM 1488 Rd. | Suite 120 #248 | Conroe, TX 77384 | justfacts.com

Email From Kasprak to Richman on 5/19/24

Hello Dr. Richman,

I am a reporter for the fact-checking website Snopes.com. I am writing an article about the claim that "10-27% of non-citizens are registered to vote," which stems from a post authored by James Agresti and which is based on your 2014 paper. You and I spoke back in 2017, when Agresti also used that same work to argue that millions of non-citizens voted for President Obama in the 2008 presidential election.

You previously wrote that your best guess was that "one quarter of non-citizens were likely registered to vote." In the report you authored as an expert witness in the Mi Familia Vota v. Hobbs Arizona voting law case, you used two separate methods that were "distinct methodologically" from the 2014 paper to conclude, in both cases, a non-citizen registration rate of less than one percent.

You state in that report that this methodology has the benefit of "dramatically" reducing "the risk that response error on the citizenship status question is leading to bias."

Here are my questions:

> Do you feel as though the discrepancy between your earlier work (which did not have this benefit) and the present work (which is orders of magnitude lower) lends credence to the critics of your 2014 paper?

To a very limited extent. The estimates from the earlier paper to which the new estimates are comparable are the ones based on validating registration plus stated registration. The new estimates are within the margin of error of those older estimates, so it isn't clear that they are in any sort of conflict.

> Do you think James Agresti's recent use of your work to argue that 10-27% of non-citizens are a reasonable use of data and your methodologies given the limitations you describe about the earlier CES surveys?
>
> Do you consider the pool of non-citizen respondents in the CES surveys to be a convenience sample? How do you know that this group is representative of that population as a whole?

Thanks in advance for your time, my deadline is EOD tomorrow (Monday 5/20).

Best,

Alex Kasprak

Email From Richman to Kasprak on 5/20/24

On Mon, May 20, 2024 at 11:53 AM Richman, Jesse T. <XXX@odu.edu> wrote:

Mr. Kasprak,

Thank you for reaching out to me. I've responded to the questions you asked below. Please feel free to follow up.

> **Do you feel as though the discrepancy between your earlier work (which did not have this benefit) and the present work (which is orders of magnitude lower) lends credence to the critics of your 2014 paper?**

It lends some credence, but with important caveats.

In terms of credence, the fact that the estimates are lower supports the possibility that there was some bias generated by response error on the citizenship status question as the critics alleged.

The caveats. The estimates from the earlier paper to which the new estimates are comparable are the ones based on validating registration plus stated registration. Those were not the numbers Trump liked to waive around, and they are not the ones emphasized in the JustFacts report. Thus, the difference is from 3.1 percent to a bit less than one percent. Furthermore, as I noted in the Arizona report, there is no evidence in the recent CCES data of the kind of inconsistent responses to the citizenship question that critics hypothesized, so there isn't direct evidence for the critics' main claim.

> **Do you think James Agresti's recent use of your work to argue that 10-27% of non-citizens are a reasonable use of data and your methodologies given the limitations you describe about the earlier CES surveys?**

I think the 10 percent figure has a foundation in my work, though it needs to be caveated more clearly to indicate the risks of being biased upwards, and I think it is absolutely inappropriate as a lower bound. The appropriate lower bound should be based upon individuals with validated registration status who said they were registered (i.e. less than one percent in 2022). I don't think the 27 percent figure is reasonable, partly because of changes in the methodology of the CES, and partly because I don't fully understand how it was calculated.

The ten percent number appears to roughly approximate what one finds if one looks at the portion of survey respondents for whom there was any indication of possible voter registration – either they said they were registered, or they had a voter file match, or both.

A concern with this measure is that this measure has a much larger risk of being biased by false positives (people who said they were registered but were not, or who had a voter file match that was in error) than the numbers in my Arizona report. But it also avoids

most risk of being biased downwards. Thus, I understand where such a number might come from, though I think any presentation of it would need to be clearly delineated in terms of the assumptions behind it: basically it's based upon assuming that anyone who has any indication in the CES survey whatsoever that they were registered to vote and a non-citizen is in fact a registered non-citizen. And to reiterate, I think it's more plausible as an upper bound than a lower bound.

Here are the survey tabulations, weighted and unweighted, for the 2022 CES non-citizen subsample. The "TS_voterstatus" variable indicates voter registration file match. The "Voter Registration Status Post" variable indicates that the respondent said they were registered to vote in the post-election survey.

```
. tabulate TS_voterstatus if cit1==2, missing

voterstatus |      Freq.     Percent        Cum.
------------+-----------------------------------
     Active |         45        7.21        7.21
          . |        579       92.79      100.00
------------+-----------------------------------
      Total |        624      100.00

. tabulate votereg_post if cit1 == 2, missing

      Voter |
 Registratio|
   n Status |
       post |      Freq.     Percent        Cum.
------------+-----------------------------------
        Yes |         32        5.13        5.13
         No |        441       70.67       75.80
 Don't know |         18        2.88       78.69
          . |        133       21.31      100.00
------------+-----------------------------------
      Total |        624      100.00

.
. tabulate TS_voterstatus if cit1==2 [aweight = commonweight_hisp], missing

voterstatus |      Freq.     Percent        Cum.
------------+-----------------------------------
     Active | 50.3365228        8.07        8.07
          . | 573.663477       91.93      100.00
------------+-----------------------------------
      Total |        624      100.00

. tabulate votereg_post if cit1 == 2 [aweight = commonweight_hisp], missing

      Voter |
 Registratio|
   n Status |
       post |      Freq.     Percent        Cum.
------------+-----------------------------------
        Yes | 39.05151723        6.26        6.26
         No | 427.678301       68.54       74.80
 Don't know | 17.2446531        2.76       77.56
          . | 140.0255286       22.44      100.00
------------+-----------------------------------
      Total |        624      100.00
```

If we combine these two together, we get the following results, weighted and unweighted.

```
. generate combinedregstatus = 0

. replace combinedregstatus = 1 if TS_voterstatus == 1
(45 real changes made)

. replace combinedregstatus = 1 if votereg_post == 1
(26 real changes made)

. tabulate combinedregstatus
```

combinedreg status	Freq.	Percent	Cum.
0	553	88.62	88.62
1	71	11.38	100.00
Total	624	100.00	

```
. tabulate combinedregstatus [aweight = commonweight_hisp]
```

combinedreg status	Freq.	Percent	Cum.
0	538.487195	86.30	86.30
1	85.5128049	13.70	100.00
Total	624	100.00	

This suggests at the top end an estimate of about 11.4 unweighted or 13.7 percent weighted of the non-citizen subsample of the CES in 2022 had some indication of registration (either they said they were registered, or they had a voter file match, or both.

The confidence interval for the weighted estimate goes up to about 16.5% and down to about 11 percent. I presume an analysis similar to this was the basis for the 10 percent figure. To reiterate, if this process is how the 10 percent figure came about, I think it's an appropriate use of the methods of the paper, though it needs to be clearly identified what assumptions are being made and why they might be biasing the estimate up, and I disagree that this constitutes the lower bound.

```
. cii proportions 624 85
```

Variable	Obs	Proportion	Std. err.	Binomial exact [95% conf. interval]	
	624	.1362179	.0137318	.1102761	.1656573

```
.
. cii proportions 624 71
```

Variable	Obs	Proportion	Std. err.	Binomial exact [95% conf. interval]	
	624	.1137821	.012712	.0899415	.1413466

Now to the 27 percent number. In the 2014 paper my coauthors and I discussed an adjusted estimate. Here was our discussion of the adjustment methodology from the 2014 paper:

"It uses the 94 (weighted) non-citizens from 2008 for whom Catalyst obtained a match to commercial and/or voter databases to estimate the portion of non-citizens who either claim to be registered when they are not (35%) or claim not to be registered when they are (18%). We then use these numbers to extrapolate for the entire sample of non-citizens in 2008 and 2010."

In my view, the adjustment strategy we applied was more appropriate in the earlier years of the data when Catalist was matching CES respondents not only with the voter file, but also with a variety of other commercial databases. I think it makes little sense in the context of the Target Smart voter file match which only identifies individuals who are registered.

If we were to apply this methodology to the 2022 data, we wouldn't have any observations in the cell for which a Targetsmart match occurred but indicated they were not registered – this was only possible in the methodology of the earlier paper because of the commercial database matching by Catalist. Thus, I'm not sure how one would even apply the methodology to the 2022 data as a result of not being able to identify the value of this cell. If one assumed that all individuals not matched to the voter file didn't register to vote, one would get estimates out of this identical to the tabulation for "TS_voterstatus" indicated above. This would give a number of 7 or 8 percent (unweighted versus weighted) rather than 27 percent. On the other hand, if one assumed that only individuals who were matched by Target Smart "counted" for the adjustment, one would get an estimate that the 'true' level of registration was 100 percent.
Thus, I'm unsure how the 27 percent number came about, I don't think it has a foundation in my methodology, and I don't think it's plausible in light of various other indicators of the frequency of non-citizen registration and voting.

Do you consider the pool of non-citizen respondents in the CES surveys to be a convenience sample? How do you know that this group is representative of that population as a whole?

Like the CES as a whole, the subsample of non-citizen respondents is not a true random sample.

In the 2014 Electoral Studies paper my coauthors and I discussed this issue. We wrote on page 151:

> "It is impossible to tell for certain whether the non citizens who responded to the survey were representative of the broader population of non-citizens, but some clues can be gained by examining education levels. Census bureau estimates (Census, 2012) suggest that the sample contains slightly more college-educated respondents (30.6 percent) than the overall foreign born population (26.8 percent), and many fewer respondents with less than a high-school education (8.3 percent versus 33.3 percent). The paucity of uneducated non-citizens in the sample would in most circumstances be expected to bias sample voting participation upward. However, given our results concerning the association between participation and education (discussed below) it may well be that the paucity of uneducated non-citizens in the CCES sample biases the turnout estimates down rather than up. We confront this issue primarily by weighting the data."

The CES, like many other polls and surveys today, uses a methodology that attempts to approximate the distribution of respondents in the population of interest without using a random sample of the population. Given the low response rates for even "random sample" data, with rare exceptions this is how most public opinion data is produced today.

Please feel free to follow up with any additional questions.

Jesse

Email From Kasprak to Richman on 5/20/24

From: Alex Kasprak <XXX@snopes.com>
Sent: Monday, May 20, 2024 3:11 PM
To: Richman, Jesse T. <XXX@odu.edu>
Subject: Re: Media Inquiry from Snopes: "10-27% registration rate for noncitizens"

Thanks for your response. I may have a few more questions in a bit, but one that jumped out to me is your equating the non-randomness of the CES as a whole with the subset of people identifying themselves as non-citizens who are registered. I understand that CES attempts to simulate a random sample by matched pairing to a theoretical random sample. While it may not be a "true" random sample, surely you are not suggesting that both the full dataset and the noncitizen subset are equally flawed from a representative standpoint, are you?

Second Email From Richman to Kasprak on 5/20/24

From: Richman, Jesse T.
Sent: Monday, May 20, 2024 4:32 PM
To: Alex Kasprak <XXX@snopes.com>

Subject: RE: Media Inquiry from Snopes: "10-27% registration rate for noncitizens"

I didn't mean to equate specifically with the subset of non-citizens who are registered. More with the subset of non-citizens. Of course I recognize differences, but I also recognize similarities. Maybe this would be easier to talk through on the phone than over email. You are welcome to call my cell phone at 757-201-5484. Anyhow, since we're on e-mail…

According to the CCES 2022 codebook:

"The sample drawn for the CES were chosen from the YouGov Panel, along with the Dynata, Disqo, Prodege, and Generation Lab panels using a six-way cross-classification (age × gender × race × education × region × sample source). All respondents who completed the pre election survey were re-invited to the post-election survey. The final set of completed pre election interviews (numbering 80,233, after quality controls were applied) were then matched to the target frame, using a weighted Euclidean distance metric conditioning on registration status × age × race × gender × education."

The differences:

Since non-citizens are not part of the target frame, to the extent that panel members who are non-citizens were not getting matched with citizens in the target frame, we're going to potentially have some of the non-citizens who responded to the panel not make it through to the final sample. This may result in a non-citizen sample that is a bit smaller than it needs to be and that looks a bit more like the citizen population. It would be nice if the CES would release a match of the non-citizen respondents to a non-citizen target frame.

The similarities:

The entire CES dataset is in fact drawn from a set of opt-in panels: "YouGov Panel, along with the Dynata, Disqo, Prodege, and Generation Lab panels." Thus, at a fundamental level it's all something approximating a quota / convenience sample. Everything on top of that is about some tweaking of who gets into the final published dataset, and about the design of weights to aim to make the opt-in panel match as closely as possible the target population.

Jesse