Memo On Incorrectly Assigned Respondents

Pia Deshpande

January 18, 2022

925 respondents in the 2020 CES were assigned to incorrect congressional districts, leading to them being shown candidate names for the wrong district. All of these respondents lived in North Carolina (inputstate = 37), and therefore only House races were affected.

More specifically, 922 respondents were assigned incorrect congressional districts in the pre-election survey, and 774 were assigned incorrect CDs in the post-election survey. In total, there were 925 respondents who were incorrectly assigned in at least one of the two survey waves.

To document this error and aide researchers in correcting their work, a new CES dataset has been uploaded which includes new fields:

1. The variables `cd_change_flag` and `cd_change_flag_post` document whether or not a respondent was assigned to the wrong congressional district in pre and post-election waves respectively.
2. The variables `cdid116_fix` and `cdid116_fix_post` then list the district this respondent should have been assigned to (their corrected congressional district) in pre and post-election waves.

The below tables show how candidates flagged for correction (FLAGGED) differed from those that did not need to be altered (UNCHANGED) on the house vote variable in the pre and post-election wave. This analysis uses the `commonweight` for the pre-election survey and `vvweight_post` for the post-election survey.

As you can see, respondents who were assigned to incorrect congressional districts were more likely to select that they were "Not sure" or indicate that they did not vote in this election — because the listed candidates were not from the congressoinal district they actually resided in.

Pre-election survey

House Vote Pre-Election	FLAGGED	FLAGGED_n	UNCHANGED	UNCHANGED_n
House Candidate 1	60%	106.8	71%	138.8
House Candidate 2	20%	35.8	24%	47.6
I'm not sure	6%	10.0	3%	5.5
I didn't vote in this election	8%	13.5	2%	4.2
Other	7%	11.6	0%	0.0

Post-election survey

House Vote Post-Election	FLAGGED	FLAGGED_n	UNCHANGED	UNCHANGED_n
House Candidate 1	37%	208.7	49%	287.7
House Candidate 2	45%	253.9	45%	268.0
I did not vote	1%	2.9	2%	13.5
I did not vote in this race	5%	27.3	3%	17.9
Not sure	7%	39.2	1%	4.6

Other	6%	34.8	0%	1.5

Researchers can find what House candidates were shown to respondents by looking up the variables `HouseCand1Name` and `HouseCand2Name` for relevant respondents. They can also find the names of candidates that should have been shown to this respondent by using the respondent's corrected district.

Here is are two frequency tables of candidate names incorrectly presented to respondents by frequency. The first table are a summary of`HouseCand1Name` and the second are a summary of `HouseCand2Name`. A blank under either candidate name field is not a mistake — some districts did not have a second candidate running. This row on either table represents the frequency with which a respondent was shown no name due to being incorrectly assigned to the wrong congressional district.

HouseCand1Name	Freq_pre	Freq_post
Kathy Manning	15%	15%
Scott Huffman	13%	12%
David Price	12%	13%
David Parker	12%	11%
David Wilson Brown	11%	11%
Deborah Ross	11%	11%
G. K. Butterfield	9%	9%
Patricia Timmons-Goodson	5%	5%
Morris Davis	4%	4%
Christopher M. Ward	3%	3%
Cynthia Wallace	3%	2%
Alma Adams	1%	1%
Daryl Farrow	1%	1%

HouseCand2Name	Freq_pre	Freq_post
Lee Haywood	15%	15%
Ted Budd	13%	12%
Robert Thomas	12%	13%
Patrick McHenry	12%	11%
Virginia Foxx	11%	11%
Alan Swain	11%	11%
Sandy Smith	9%	9%
Richard Hudson	5%	5%
Madison Cawthorn	4%	4%
David Rouzer	3%	3%
Dan Bishop	3%	2%
	1%	1%
Greg Murphy	1%	1%

Please direct any further questions about this error to Pia Deshpande (pia.deshpande@tufts.edu) and Brian Schaffner (brian.schaffner@tufts.edu).