Cooperative Congressional Election Study

Guide for
2012 Vote Validation Data

prepared by *Ivelisse Cuevas-Molina, Research Assistant*
University of Massachusetts Amherst

Stephen Ansolabehere, PI
Harvard University

Brian Schaffner, Co-PI
University of Massachusetts Amherst

July 16, 2017

1 Introduction

The Cooperative Congressional Election Study (CCES) is a biennial online large sample survey conducted by YouGov that includes both a pre-election and post-election wave. Every iteration of the CCES since 2008 includes vote validation conducted in cooperation with Catalist; a progressive voter file firm that matches CCES respondents to their registration and voting records. This guide will describe the validation variables included in the 2012 CCES, how to interpret these variables and how to use them in conjunction with self-reports of registration and turnout. The 2012 CCES has a total sample of 54,535 respondents.

For details on the Catalist matching process of the CCES please see Ansolabehere and Hersh (2010).

2 Vote Validation Variables

Vote validation for the 2012 CCES includes seven (7) vote validation variables, which are listed in this section: (1) catalist_match, (2) voter_status, (3) e2012g, (4) e2012presp, (5) e2012congp, (6) presprim_pty, and (7) congprim_pty.

The "catalist_match" variable indicates whether a respondent within the 2012 CCES was matched to the Catalist voter file; 82.92%[1] of respondents were matched to a record. Only matched respondents have a verified record of turnout or non-participation in the 2012 General Election. Non-matched respondents appear as missing values in all of the vote validation variables, but the "catalist_match" variable. Section 4 of this guide will discuss how to interpret non-matched respondents in light of vote validation accuracy.

Validated voter registration status is measured with the "voter_status" variable. Note that this is the individuals voter registration status at the time of the match, which happens in June of the year following the election. Most matched respondents in the 2012 CCES have active voter registration records (88.92%[2]). Some matched respondents had been dropped from voter rolls, had inactive voter registration, are listed as "multipleAppe" or having multiple appearances. The multiple appearances code means that the individual was located by Catalist in a one state but was registered to vote in a dif- ferent state. This typically happens because somebody has filed a change of address form with the post office but has not yet changed their voter registration to a new state. Finally, some matched respondents are confirmed as unregistered.

The "e2012g" variable measures mode of participation in the 2012 General Election. All respondents in this variable other than those in the "MatchedNoVote" category are validated voters, and missing values correspond to non-matched respondents. Respondents in the "MatchedNoVote" category are matched respondents who are confirmed non-voters. The "e2012presp" measures mode of participation in the 2012 presidential primary election, and "presprim_pty" indicates whether those who participated in the presidential primaries voted in the Democratic or Republican primaries. Similarly, the "e2012congp" variable measures mode of participation in the 2012 congressional primary elections, and "congprim_pty" indicates whether participation occurred in a Democratic or Republican primary. However, not all states record which partys primary a voter participated in, so there are values of "presprim_pty and "e2012congprim_pty" for only a subset of people who have a vote record on "e2012prep and "e2012congp."

catalist_match	Successful match to Catalist	
0	No	9,314
1	Yes	45,221

[1]This quantity represents the unweighted percent of 2012 CCES respondents who were matched to Catalist.

[2]This quantity represents the unweighted percent of matched respondents in the 2012 CCES with an active registration record.

voter_status Catalist - voter status

active	40,210
dropped	897
inactive	513
multipleAppe	246
unregistered	3,355

e2012g Catalist - vote history general election

Absentee	6,101
Early	5,206
Mail	2,464
MatchedNoVote	8,891
Polling	11,810
UnknownMethod 10,821	

e2012presp Catalist - vote history presidential primary

Absentee	606
Early	415
Polling	2,210
Unknown	1,571

e2012congp Catalist - vote history congressional primary

Absentee	2,539
Early	1,935
Mail	1,814
Polling	6,664
Unknown	3,914

presprim_pty Catalist - presidential primary party

Democratic	413
Republican	1,537

congprim_pty Catalist - congressional primary party

Democratic	4,082
Republican	5,325

3 Registration and Turnout Self-Report Variables

CCES respondents self-report their registration in both the pre- and post-election waves of the survey. The "votereg" and "votereg_post" variables measure self-reported registration, but the most consequential of these two variables is the post-election item. Only those respondents who self-report "Yes" to the "votereg_post" question are asked the turnout self-report question, namely "CC401".

The turnout self-report question, "CC401", is asked in the post-election wave of the CCES, and has five response options. Four of those response options indicate non-participation (1 through 4) and only one response option indicates participation in the 2012 General Election (5-"I definitely voted in the General Election on November 6.").

Those respondents who indicate that they are not registered to vote by answering "No" to the "votereg_post" question should be considered *self-reported non-voters* for all purposes of analysis of self-reported turnout among 2012 CCES respondents. Researchers should combine self-reported non-registered respondents and those who selected categories 1 through 4 in the turnout self-report question in order to identify all *self-reported non-voters*. Only those who selected category 5 of "CC401" are *self-reported voters*.

Please note that all respondents from North Dakota are asked the turnout self-report question no matter their response to the registration self-report questions because this state does not require registration for participation.

votereg	Voter Registration Status	
1	Yes	50,306
2	No	3,704
3	Don't Know	525

votereg_post	Voter Registration Status	
1	Yes	42,397
2	No	2,449
3	Don't Know	172
.	NA	9,517

CC401	Turnout	
1	I did not vote in the election this November.	1,248
2	I thought about voting this time - but didn't.	473
3	I usually vote, but didn't this time.	606
4	I attempted to vote but did not or could not.	550
5	I definitely voted in the General Election on November 6.	39,675
.	NA	11,983

4 Using Vote Validation with Self-Reports

There are three possible ways to measure turnout in the 2012 CCES using the validation variables. Two use only the "e2012g" vote validation variable while the third uses this variable in conjunction with self-reported registration (votereg_post) and self-reported turnout (CC401). The STATA code below represents all three specifications.

STATA Code for 2012 CCES Validated Turnout

```
*2012 CCES Validated Turnout
use "~/CCES12_Common_VV-2.dta", clear

* including all non-matched as nonvoters N=54,535
gen voter=1 if e2012g !="MatchedNoVote" & e2012g !=""
replace voter=0 if e2012g =="MatchedNoVote" | e2012g ==""

* excluding all non matched N=45,221
gen voter2=1 if e2012g !="MatchedNoVote" & e2012g !=""
replace voter2=0 if e2012g=="MatchedNoVote"

* including only non matched respondents
* who are self-reported nonvoters N=46,956
gen voter3=1 if e2012g !="MatchedNoVote" & e2012g !=""
replace voter3=0 if e2012g=="MatchedNoVote"
replace voter3=0 if catalist_match==0 & votereg_post ==2
replace voter3=0 if catalist_match==0 & CC401<5
```

1. **Un-matched as non-voters.** The first specification defines *voters* as respondents with a validated voting record no matter their mode of participation, and defines *non-voters* as both matched non-voters and non-matched respondents. This specification retains the integrity of the full CCES sample, no missing values are created. The justification for the second approach is the fact that the most common reason that Catalist will not have a record for an individual is because that individual is not registered to vote. Indeed, rates of self-reported non-registration and non-voting are much higher among un-matched respondents than among those for whom there is a match.

2. **Only matched non-voters as non-voters.** The second specification defines *non-voters* as only matched non-voters. This specification reduces the CCES sample and results in validated turnout estimates that are larger than those in the first specification. However, this specification increases the level of certainty in the identification of non-voters in the CCES, because there could possibly be actual voters among non-matched respondents.

3. **Matched non-voters and self-reported non-voters as non-voters.** The third specification defines *non-voters* as (1) matched non-voters, (2) non-matched respondents who reported not being registered to vote in the "votereg_post" question, and (3) non-matched respondents who are self-reported non-voters in the "CC401" question. This definition excludes non-matched respondents who are self-reported voters (these individuals would be coded as missing). This definition assumes that self-reported non-voters are honest about their non-participation because there is no incentive to go against the democratic norm of participation.

5 Vote Validation Accuracy

How do these different approcahes compare in terms of estimating aggregate turnout rates? Here we provide an accuracy analysis comparing state-by-state turnout estimates from Catalist vote

validation of the 2012 CCES with VAP[3] turnout estimates from the United States Elections Project (USEP)[4].

Figures 1 and 2 present turnout accuracy plots for Catalist vote validation of the 2012 CCES. Estimates of CCES validated turnout rates from the 50 states of the Union and the District of Columbia are plotted with 95% confidence intervals. The 45 degree line in each plot represents VAP turnout or "actual turnout"; the closer the estimates are to "actual turnout" the more accurate the validated turnout estimate. Figure 1 shows that validated turnout estimates that include non-matched respondents (method 1) are closer to VAP turnout. Estimates in Figure 2 show that excluding non-matched respondents from estimates of validated turnout result in a much larger overestimation of turnout (method 2). Again, this is beacuse many non-voters do not have records in the Catalist database.

Though the plots in Figures 1 and 2 are informative, Mean Square Error (MSE) statistics can provide more a precise measurement of the discrepancy between CCES validated turnout and USEP VAP turnout. Interpretation of this statistic is quite simple. The closer to zero the better, meaning there is a smaller distance between the estimates and the true value. The MSE for CCES validated turnout including non-matched respondents is 0.0070 and the MSE for validated turnout excluding non-matched respondents is 0.0489. The latter statistic is larger than the former, suggesting that defining non-matched respondents in the CCES as *non-voters* is more accurate than excluding them from analysis.

[3] Voting Age Population

[4] McDonald, Michael P. 2017. "2012 November General Election Turnout Rates" United States Elections Project. April 11,2017. http://www.electproject.org/2012g

Figure 1. Turnout Accuracy Including Non-Matched Respondents

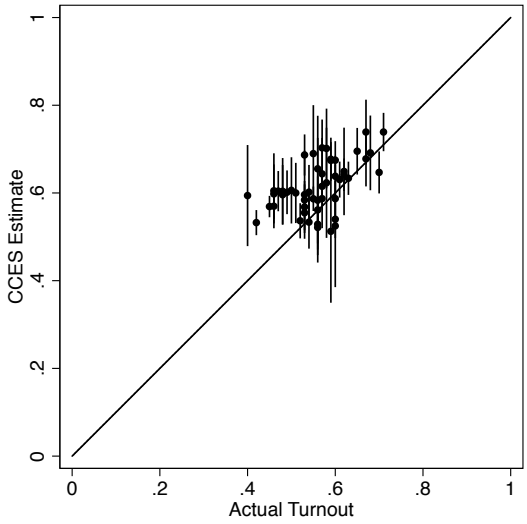

Figure 2. Turnout Accuracy Excluding Non-Matched Respondents

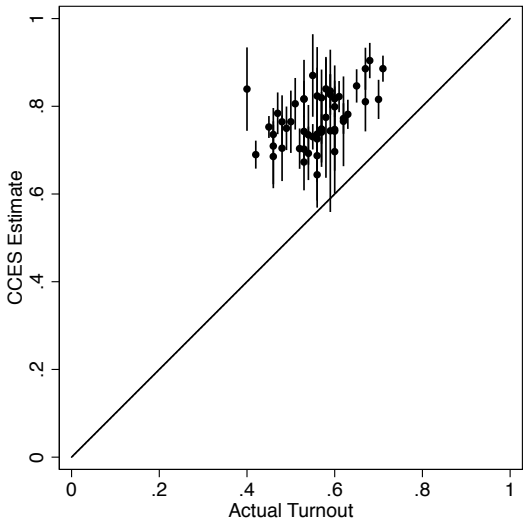

Figure 3 shows that the third method (treating matched non-voters and un-matched self-reported non-voters are non-voters) produces validated turnout estimates that are worse on average that the first method, but better that method 2. The MSE fot this third specification is 0.0306 .

Still, though it is likely that most non-matched respondents are actual non-voters the could be actual voters among them. it is the discretion of the researcher how to treat these non-matched respondents when studying turnout among CCES respondents.
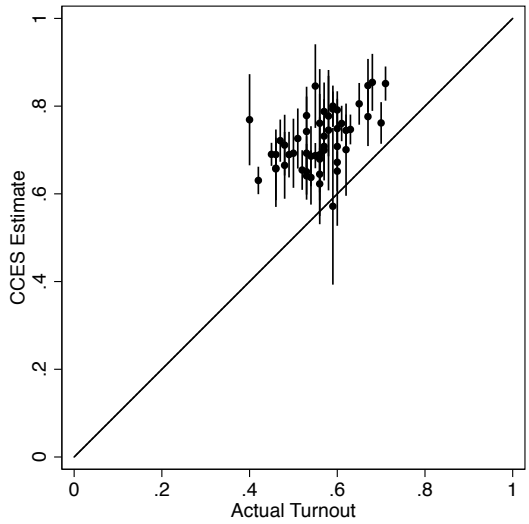
Figure 3. Turnout Accuracy Including Matched Non-voters and Non-Matched Self-Reported Non-voters