VOTED IN THE LAST ELECTIONS - General Target Variable Report (GVR)

1. General Information

The target variable T_VOTED measures respondents' reported participation in the most recent (last) parliamentary or presidential elections in their country. T_VOTED is a nominal variable that takes values from 1 to 7 (see Table 1.1). Values 1 and 2 indicate that respondents have voted, although some report having casted a blank or invalid ballot (value 2). Values 3 to 7 capture non-voting and distinguish between various reasons for not voting when such information is available in the source data.

The target variable T_VOTED is accompanied by the harmonization control variable C_VOTED_TYPE (see Table 1.1). This measure indicates if target values on T_VOTED refer to respondents' participation in the last parliamentary elections (value 1), in the last presidential election (value 2), in either or both, since the last parliamentary and presidential elections took place simultaneously (value 3), or if the type of elections cannot be determined using the source documentation or external information about the country's type of political system (value 4).[1]

The target variable report for T_VOTED is accompanied by the following Excel documents:

- Detailed Variable Report (DVR): T_VOTED_DVR_SDR2.xlsx. DVR Excel files in SDR2 systemize all information about source variables that were used for harmonization into a given target variable of the SDR2 database;
- Crosswalk Table (CWT): T_VOTED_CWT_SDR2.xlsx. CWT Excel files in SDR2 contain details about mapping of source values to target values.

Table 1.1 VOTED IN THE LAST ELECTIONS: Description of the target, source, and control variables

	Variable description	**Variable name**	**Variable values[a]**
Target variable	Respondent's declared behavior in the last national elections	T_VOTED	1 = Voted 2 = Voted but casted a blank/invalid ballot

[1] Rarely, a survey asks separate questions about participation in the last presidential and in the last parliamentary elections without specifying which election was the most recent. In such instances, we prioritize the source information that corresponds to the political system of the country in question (e.g. for a country with a presidential system, we harmonize the source information about voting in the last presidential election). We detail these decisions in Sections 3.2.2 and 3.2.3 of this report. .

			3 = Did not vote 4 = Did not vote: not eligible (too young, non-citizen status) 5 = Did not vote: not registered 6 = Did not vote due to external pressure 7 = Did not vote inferred from other answer categories (e.g. don't remember, NAP)
Source variables			See: T_VOTED_DVR_SDR2.xlsx and T_VOTED_CWT_SDR2.xlsx
Control variables	Type of last election that respondents are asked about	C_VOTED_TYPE	1 = Last parliamentary elections 2 = Last presidential election 3 = Last parliamentary and presidential elections took place on the same day 4= Not identified if parliamentary/presidential elections

[a] Missing values are assigned according to the SDR2 missing codes schema, provided in the Appendix.

2. Survey Projects

Source variables that we used for T_VOTED appear in 21 international survey projects: ABS, AFB, AMB, ARB, ASES, CB, CDCEE, CNEP, EB, EQLS, ESS, ISJP, ISSP, LB, LITS, NEB, PA1, PA2, PPE7N, VPCPCE, WVS, 102 waves and 1651 national surveys. The data cover 142 countries and years from 1967 to 2017.

3. General Rules and Procedures

3.1. Source data description

To construct the target variable T_VOTED, we use source items about respondents' participation in the most recent parliamentary or presidential elections in their country. The main differences in source measures of respondents' reported voting behavior can be summed up as follows:

- Source questions ask either about (i) taking part in elections in general, or (ii) taking part in a particular type of election, or (iii) the name of a specific party or candidate the

respondent voted for. Variation in wording notwithstanding, the most common formulations of questions are as follows:

Did you vote in the last/recent/previous national/general **parliamentary** elections?
Who/which party did you vote for in the last **parliamentary** elections?
Did you vote in the last/recent/previous national/general **presidential** elections?
Who/which party did you vote for in the last **presidential** election?

- The large majority of 'voting' source variables that ask about participation (rather than names of party or candidates) provide "Yes," "No" answer options, accompanied by various types of missing values. However, in rare instances the answer option "did not vote" is combined with other possibilities, e.g., "Can't remember" (CNEP/4 `H.VoteWhichRecent` for Spain), "Not applicable" (CNEP/4 `H.VoteWhichRecent` for Indonesia), and "NAV" (ISSP/1993 `v308` for Australia). See also Section 3.2.3, discussion of a target value 7.

- Source measures differ in the amount of detail they provide about "yes" (voted) or "no" (did not vote) answers. Thus, we learn that some respondents who voted, casted empty or invalid ballots. We also learn about a range of reasons for why respondents did not vote. Whenever such information is available in the source data, we retain it in the target variable T_VOTED (see Table 1.1).

3.2. Rules of transformation of source variables into target variable

3.2.1. For ex-post harmonization we select only those source variables that refer to respondents' reported participation in the most recent parliamentary or presidential elections.

To do so, we first carried out a keyword-based search in the SDR 2.0 "Cotton file: Cumulative List of Variables in the Surveys of the SDR Database"[2] that included words such as: "party," "election," "vote," "voting," "parliament(ary)," "president(ial)," "general," "national," "federal," "legislative," "legislator." The search was comprehensive, i.e., it included variable's name, variable's label, and values.

Second, we removed source variables that did not measure reported participation in last elections. In some cases, respondents were asked about their voting history, what party they would vote for if elections were organized in the near future (e.g. "next Sunday"), or which party they felt closest to at the moment of the interview.

[2] Powałko, Przemek, 2019, "SDR 2.0 Cotton File: Cumulative List of Variables in the Surveys of the SDR Database", https://doi.org/10.7910/DVN/6QBGNF, Harvard Dataverse, V2

Third, we removed source variables that measure reported participation in other types of elections than parliamentary or presidential. This included: votes casted in local and regional elections, elections to the European Parliament, and referenda.

After completing these steps, we retained two types of 'voting' source variables: items that asked about the very act of voting (i.e. without additional specifications like party preference) and items that contained specific information on party or candidate selection next to information on whether a person took part in elections or not.

3.2.2. When a source data file provides more than one variable about respondent's participation in last elections, we select only one source measure for ex-post harmonization. The following rules inform our selection decisions:

- When source variables provide the exact dates of the last elections (e.g., via question wording or variable labels), we select the variable measuring respondents' participation in the most recent election. Such situations occur when a data file provides measures of participation in both parliamentary and presidential elections, and/or when it provides measures of participation in both rounds of a presidential election (e.g. for both Russia in CDCEE/1-2 and Ukraine in NEB/7 we harmonized the variables measuring respondents' participation in the second round of presidential election).
- If there are two separate questions (as in AMB for El Salvador, as well as LB/2000, LITS/2 and LITS/3) referring to the last parliamentary and presidential elections, and the source documentation does not specify dates, we select the source variable that measures participation in the country's more powerful political institution (e.g. we take voting in presidential elections for countries that have presidential and semi-presidential political systems, and voting in parliamentary elections for countries that have parliamentary systems). The division of countries into types of political systems is based on the open-access sources.[3]

3.2.3. When the answer option "did not vote" is not available in the source data as a separate answer category but can be inferred from other answer categories, we assign it the target value 7. Specifically, the target value 7 captures the cases when a source value combines "did not vote" answers with other possibilities (i.a. different types of missing values), for example "can't remember" (CNEP/4 Spain) and "not applicable" (CNEP/4 United States). We also assign the target value 7 when there is one source category for those who "did not vote" and those who are "not eligible to vote" (e.g. ISSP/1993).

[3] .cia.gov/the-world-factbook/field/government-type

cs.mcgill.ca/~rwest/wikispeedia/wpcd/wp/l/List_of_countries_by_system_of_government.htm

In a few cases, we inferred that "did not vote" responses are indirectly captured in other answer categories. It usually happens in source variables that capture voting for particular parties or candidates and none of their labels directly mention "did not vote" answer option. In such instances, we assume that some answer categories - "none" (e.g. ISSP/1987 `v106` for Australia) and "no preference" (e.g. EB/11 `v10`) - Missing values codes - "Inapplicable" ("NAV," "INAP" and "NAP") (e.g. CNEP/4 Indonesia) and "no answer" ("NA") (e.g. ISSP/1993 `v316`) as well as their combinations (e.g. EB/24 `v75` for Spain and Portugal) include the information that a respondent did not vote.

Missing values and different situations that warrant to be treated as missing data are coded according to the SDR2 missing codes schema, provided in Table A.1 in the Appendix.

3.3. Methodological variables that accompany T_VOTED

The target variable T_VOTED is accompanied by the harmonization control variable C_VOTED_TYPE (see Table 1.1).

C_VOTED_TYPE indicates whether 'voting' source variables refer to respondents' participation in the last parliamentary elections (value 1), in the last presidential election (value 2), in either or both, since the last parliamentary and presidential elections took place on the same day (value 3), or if it is not possible to determine the type of elections on the basis of available source documentation or external information about the country's type of political system (value 4).

To do so, we rely on a wide array of source data information and a variety of procedures.

First, we use external classifications of countries' political systems to identify countries that did not have direct presidential elections in 1966-2017.[4] These are the following:

> Angola, Anguilla, Antigua and Barbuda, Australia, Bahrain, Bangladesh, Barbados, Belgium, Belize, Bermuda, Bhutan, Botswana, Cambodia, Canada, Cayman, Czech Republic (until 2008 included), Cook Islands, Denmark, Estonia, Ethiopia, Falkland Islands, Faroe Islands, Germany, Gibraltar, Greece, Guyana, Hungary, India, Iraq, Israel, Italy, Jamaica, Japan, Jordan, Kuwait, Latvia, Lebanon, Libya, Liechtenstein, Luxembourg, Malaysia, Malta, Mauritius, Monaco, Morocco, Myanmar, Netherlands, New Zealand, Norway, Oman, Pakistan, Papua New Guinea, Qatar, Samoa, San Marino, Solomon, Islands, South, Africa, Spain, Suriname, Sweden, Switzerland, Thailand, Tonga, Trinidad and Tobago, Turkey (until 2007 included), United Kingdom, Vietnam.

[4] International Foundation for Electoral Systems. "ElectionGuide data": https://www.electionguide.org/ (Accessed January 27, 2023)

This step allows us to code as "parliamentary" all source variables referring to countries that do not hold direct presidential elections. This step is also important because some wave-level source variables have misleading labels: the label refers, generally, to voting in presidential elections, but some countries within that wave do not have such elections. For example:

> LITS/3 `q419c` (variable label "Voted in the most recent presidential elections") referring to Germany, Greece, Hungary, Italy, Latvia
>
> AMB/2006 `vb2` (question: "¿Votó Ud. en las últimas elecciones presidenciales?") referring to Canada, Guyana, and Jamaica
>
> AMB/2008 `vb2` (question: "Did you vote in the last presidential elections of (year of last presidential elections)?") referring to Belize, Jamaica, and Canada

In such cases, the harmonization control C_VOTED_TYPE takes the value 1 ("parliamentary") for respondents who live in countries that do not have presidential elections, and the value 2 ("presidential") for respondents who live in countries where the citizenry elects the president. To illustrate,

> ASES `V0196` (question "Q410 Which party did you vote in PRESIDENT election?(Japan") - since Japan does not have direct presidential election, we classified it as "parliamentary"
>
> ASES `V0197` (question: "Q410 Which party did you vote in PRESIDENT election?(S.Kor") types of questions was classified as "presidential"

If the source documentation provides the dates of the last elections that respondents were asked about, the election type is identified on the basis of open-source information, e.g.

> AFB/3 `q30`: "Note: Last national election was in 2000 in Senegal and Tanzania; 2001 in Zambia; 2002 in Cape Verde, Kenya, Lesotho, Madagascar, Mali, and Uganda; 2003 in Benin and Nigeria, 2004 in Botswana, Ghana, Malawi, Mozambique, Namibia, and South Africa; and 2005 in Zimbabwe" => Since 2000 elections in Senegal were presidential elections, for respondents answering AFB/3 `q30`, C_VOTED_TYPE takes the value 2, "presidential."

C_VOTED_TYPE takes the value 3 when a source data file provides a single variable about respondents' participation in most recent elections, but the question refers to a situation when both parliamentary and presidential elections took place on the same day.

> To illustrate, in ISSP/2011, the question asked in Chile was as follows: "Are you registered to vote? (If yes) Did you vote on the last presidential and parliamentary elections of 2009?"
>
> Therefore, C_VOTED_TYPE is coded with the value 3.

Sometimes, neither checks of the source information, that is, source variable and value labels (e.g. reference to party names or to names of presidential candidates) and question wording, nor external information about the country's type of political system, allow us to identify whether the 'voting' source variable refers to the last parliamentary or the last presidential election. In such instances, the harmonization control variable C_VOTED_TYPE takes the value 4.[5]

4. Special Cases

- Projects that do not include questions about voting in last elections in their source files although, according to information provided in their documentation, they should, are: EB 63.4, 64.4, 66.1, 68.1, 73.4, 74.1, and 76.2, and NBB.
- EB/35: for the source variable `v237`, the question wording (in the English questionnaire and the variable label) indicates that it measures respondents' past behavior (i.e. retrospective voting), yet the variable's value labels use the term "would." We decided to rely on the variable label and the questionnaire, and used this variable for ex-post harmonization into T_VOTED.
- ISSP/1987 `v106`, ISSP/1990 `v86`, ISSP/1991 `v104`, ISSP/1993 `v308`: In Australia where voting has been compulsory since 1924 in instances where there is no "did not vote" option or "did not vote" answer is mixed with other categories, we inferred that "None," "None, no party" and "No preference" stand for 7, i.e. "Did not vote inferred from other answer categories." We, therefore, assumed that those respondents that chose these categories, did not vote at all and not just cast an invalid ballot.
- EB/11: depending on the country, different source codes stand for value "7." We inferred that for Germany, Luxemburg, and the Netherlands source value 99 ("DK; NA" in codebook) implies "did not vote" whereas for the remaining countries, apart from the Great Britain and Ireland which are coded as missing ("QNA"), both 99 and 96 imply that a respondent did not vote.

[5] In principle, it could be possible to identify which election respondents were asked about by cross-checking fieldwork dates with election events in a given country. Such an investigation, however, goes beyond the scope of our target variable, T_VOTED, whose purpose is to measure reported participation in the most recent general election.

Appendix: Codes for missing values in SDR2

In the SDR database v.2 we identify different situations that warrant to be treated as missing data. Table A.1 lists all SDR2 missing value codes:

Table A.1. Codes for missing values in SDR2

SDR tag [a]	SPSS (STATA) codes	Label
		Standardized source codes for missing values
DK	-1 (.a)	Don't know
NA	-2 (.b)	No answer
REF	-3 (.c)	Refusal
DU	-4 (.d)	Don't understand the question
DNR	-5 (.e)	Any combination of DK, NA, REF, DU
INAP	-6 (.f)	Inapplicable
NEC	-7 (.g)	Not elsewhere classified
		SDR created codes for missing values
UNFIT	-8 (.h)	Source value does not fit to target
ERR	-9 (.i)	Errors in source data and undocumented source values
COMBI	-10 (.j)	Different missing codes on multiple sources taken for a target
CINAP	-11 (.k)	For control variables only: inapplicable
INSUF	-12 (.l)	For survey: Insufficiently defined response categories
QNA	-20 (.t)	For survey: Question not available

[a] Abbreviations for the labels corresponding to the SDR2 codes for missing values. These tags are used in the Crosswalk Table (CWT) files (Excel) that accompany documentation of SDR2 target variables.

In exceptional situations when codes for missing data listed in Table A.1 cannot be used, we apply a system missing <null> value.