PARTICIPATION IN DEMONSTRATIONS - General Target Variable Report (GVR)

1. General Information

The target variable T_DEMONST measures respondents' reported participation in demonstrations or protests in the past, or their willingness to engage in these activities in the future. The measure is nominal (see Table 1.1). It takes the value 1 whenever respondents report that they did not demonstrate or protest in the past (in the 'yes/no' question). It takes the value 2 when respondents indicate unwillingness to engage in demonstrations or protests in the future (in the 'yes/would/would not' question). It takes the value 3 when respondents state the willingness to participate in demonstrations or protests in the future (in the 'yes/would/would not' question). Whenever respondents confirm that they demonstrated or protested in the past, T_DEMONST takes the values 4 (in the 'yes/would/would not' question) and 5 (in the 'yes/no' question).

The target variable T_DEMONST is accompanied by six harmonization control variables (see Table 1.1 and Section 3.3) specifying properties of the source items, such as: (a) time span for participating in demonstrations in the past (C_DEMONST_YEARS), (b) source item asks about other activities next to demonstrations, such as marches, protests or sits-in (C_DEMONST_EXTENDED), (c) source item asks about the illegal character of the demonstration (C_DEMONST_ILLEGAL), (d) source surveys contains more than one question on participation in demonstrations and we use them to construct the target (C_DEMONST_SET), (e) the source item specifies the demonstration's cause/concern (C_DEMONST_ISSUE), and (d) source item wording explicitly mentions "protest" (C_DEMONST_PROTEST).

We also provide the quality control variable for processing error types, QP_DEMONST. It measures the occurrence of processing errors on source variables used to construct T_DEMONST.

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

- The Detailed Variable Report (DVR) T_ DEMONST _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;
- The Crosswalk Table (CWT): T_ DEMONST _CWT_SDR2.xlsx. CWT Excel files in SDR2 contain details about mapping of source values to target values.

Table 1.1. PARTICIPATION IN DEMONSTRATIONS: Description of the target, source, and control variables

	Variable description	Variable name	Variable values[a]
Target variable	Participation in demonstrations and willingness to demonstrate	T_DEMONST	1 = No (question type: yes/no) 2 = Would not (question type: yes/would/would not)

			3 = Would (question type: yes/would/would not) 4 = Yes (question type: yes/would/would not) 5 = Yes (question type: yes/no)
Source variables			See: T_DEMONST_DVR_SDR2.xlsx T_ DEMONST_CWT_SDR2.xlsx
Control variables	Participation in demonstration: time frame	C_DEMONST_YEARS	1 = Last year 2 = Last 2 years 3 = Last 3 years 5 = Last 5 years 8 = Last 8 years 10 = Last 10 years 11 = Ever
	Source item asks about participation in demonstrations and other forms of protests (e.g., march or sit-in)	C_DEMONST_EXTENDED	0 = Restricted to demonstration 1 = Demonstration and other actions
	Source item asks about participation in illegal demonstration	C_DEMONST_ILLEGAL	0 = Otherwise 1 = Illegal demonstration
	Two or more source items used to construct the target variable	C_DEMONST_SET	0 = Single item 1 = Set of items on different time frames for participation 2 = Set of items on different reasons for participation 3 = Set of items on legal or illegal participation
	Source item specifies demonstration cause/concern (e.g., demonstration against environmentally harmful policy; protests during the Arab Spring, demonstration to solve problems in a neighborhood)	C_DEMONST_ISSUE	0 = Issue not specified 1 = Environment 2 = Arab Spring 3 = Neighbourhood
	Source item wording explicitly mentions “protest”	C_DEMONST_PROTEST	0 = Protest not mentioned 1 = Protest mentioned

Quality control: Processing error types (realized values)	QP_DEMONST	0 = No processing errors 3 = Contradictory variable values 5 = Lack of variable value labels 10 = Contradictory variable values & lack of variable value labels

[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_DEMONST appear in 19 international survey projects: ABS, AFB, AMB, ARB, ASES, CNEP, EB, EQLS, ESS, EVS, ISJP, ISSP, LB, LITS, PA1, PA2, PPE7N, VPCPCE, WVS, 79 waves and 1612 national surveys. The data cover 153 countries and years from 1971 to 2017.

3. General Rules and Procedures

3.1. Source data description

Source items exhibit some variation in terms of answer options and value labels. Some surveys only ask about the fact of participation in demonstrations in the past, providing binary 'yes', 'no' or 'mentioned', 'not mentioned' answer options. Others enable respondents to mark whether they participated in demonstrations in the past, or would / might or would not participate in demonstrations in the future.

We mainly rely on the English language questionnaires and codebooks describing the source survey data.

To construct the target variable T_DEMONST, we use these two types of source questions on participation in demonstrations:

(1) questions about the past behavior (i.e., reported fact), which we label as the "yes/no" question type.

The typical "yes/no" question about participation in demonstrations in the past is: *During the last 12 months, have you taken part in a lawful public demonstration?* (ESS/8).

(2) questions about past behavior and future intent (i.e., reported potential), which we label as the "yes/would/would not" question type.

The typical "yes/would/would not" question about both participation in demonstrations in the past and the willingness to sign a petition in the future is: *Please indicate whether you took part in a demonstration in the past year, whether you have done it in the more distant*

past, whether you have not done it but might do it or have not done it and would never, under any circumstances, do it? (ISSP/2014).

3.2. Rules of transformation of source variables into target variable

For ex-post harmonization, we select source variables that refer to respondents' participation in demonstrations or protests in the past, or a willingness to engage in demonstrations or protests in the future. We do not harmonize questions concerning other forms of collective action, such as riots, blocking traffic, or street violence. Neither do we take survey questions about respondents' attitudes toward demonstrations as a form of expressing opinion in a democratic society, or their fear of taking part in demonstrations.

If respondents did not participate in demonstrations or protests in the past (source questions about actual participation, which we label 'yes/no' question type) we assign the value 1; if respondents would not participate in demonstrations or protests in the future (source questions that combine information about past activity and information about willingness to participate in the future, which we label 'yes/would/would not' question type) we assign the value 2; if respondents would demonstrate or protest in the future ('yes/would/would not' question type) we assign the value 3; if respondents report that they demonstrated or protested in the past while answer options also included potential behavior ('yes/would/would not' question type) we assign the value 4; if respondents reported that they demonstrated or protested in the past ('yes/no' question type) we assign value 5 (see Table 1.1).

If a given survey "divides" the question about demonstration into a subset of questions, we use the full subset to create our target, flagging these instances with the harmonization control variable `C_DEMONST_SET`. For example there are surveys where respondents are asked about participation in lawful demonstrations (first question) and unlawful demonstrations (second question). While recoding the set of questions, we used the rule that the strongest answer wins, meaning that if the respondent answered "yes" to participation in lawful demonstrations, but "no" to participation in unlawful ones, we code it as "yes," the responded participated in demonstrations. Table 3.2.1 provides an example of recoding a set of two questions.

Table 3.2.1 Rules of recoding a set of two questions on participating in demonstrations for the same survey/country/year

1st answer	2nd answer	Recoded
Yes	Yes	Yes
No	Yes	Yes
Don't know	Yes	Yes
Missing	Yes	Yes
No	No	No
Don't know	No	No
Missing	No	No
Don't know	Don't know	Don't know
Don't know	Missing	Don't know
Missing	Missing	Missing

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_DEMONST

The target variable T_DEMONST is accompanied by six harmonization control variables and one quality control variable (see Table 1.1).

C_DEMONST_YEARS specifies the time span for participating in demonstrations in the past. It takes the value 1 when respondents were asked about participation in demonstrations in the last year / 12 months, for example: *Over the last 12 months, have you attended a protest or demonstration?* (EQLS/1-3). It takes values 2, 3, 5, 8, or 10 when source questions asked about participation in demonstrations in the last two, three, five, eight, and ten years, correspondingly. The control variable takes the value 11 when respondents were asked if they EVER participated in demonstrations, for example: *Have you ever participated in a public demonstration or protest?* (AMB/2008). Note that no source questions in the datasets we downloaded for harmonization ask about having demonstrated/protested in the last four, six, seven or nine years.

The majority of source variables that we harmonized into T_DEMONST include the time span in the question wording and thus, the harmonization control C_DEMONST_YEARS is generally measured at the level of national surveys. However, there are two exceptions to this rule, when C_DEMONST_YEARS also stores individual-level information. First, in the 2004 and 2014 waves of ISSP, the time span appears as a response option, not in the question wording: *Not done, never do; Not done, might do; In more distant past done; In the past year done*.

Second, the 2006 and 2008 waves of AMB ask all respondents two questions , each with a different time span: having *ever* participated in a demonstration, and having participated *in the last year*, respectively (we take both, as flagged in the harmonization control C_DEMONST_SET) .

Although demonstration/protest items in ISSP 2004 and 2014, and AMB 2006 and 2008 are exceptional situations, we harmonize them following the "strongest answer wins" rule, as illustrated in Table 3.3.1 (see T_DEMONST_CWT_SDR2.xlsx for more details) Thus, C_DEMONST_YEARS also contains respondent-level information and cannot be used directly as a filtering variable for the "last year" or "ever" time spans. Filtering requires additional recodes, depending if a user would like to select only demonstrated "last year," or is interested in demonstrated "ever" (see Table 3.3.1 for two examples and T_DEMONST_CWT_SDR2.xlsx for more details).

Table 3.3.1. Examples of coding C_DEMONST_YEARS in special cases when time span is coded at the individual level; more in T_DEMONST_CWT_SDR2.xlsx.

C_DEMONST_YEARS: Control variable labels	**Control values**	**Source values**	**Source labels**	**Target values**	**T_DEMONST SDR target labels**
ISSP/2014					
CINAP*	-11 (.k)	9	NA	-2 (.b)	NA*
CINAP*	-11 (.k)	8	DK	-1 (.a)	DK*
„ever"	11	4	Not done, never do	2	Would not (qt: yes/would/would not)
„ever"	11	3	Not done, might do	3	Would (qt: yes/would/would not)
„ever"	11	2	In more distant past done	4	Yes (qt: yes/would/would not)
"last year"	1	1	In the past year done	4	Yes (qt: yes/would/would not)
AMB/2008		**prot1 (ever) + prot2 (last year)**			
CINAP*	-11 (.k)	8+9	DK/DR + N/A	-10 (.j)	COMBI*
„ever"	11	3+9	Never + N/A	1	No (qt: yes/no)
„ever"	11	1+3	Sometimes + Never	5	Yes (qt: yes/no)
„ever"	11	1+8	Sometimes + DK	5	Yes (qt: yes/no)
„ever"	11	2+3	Almost never + Never	5	Yes (qt: yes/no)
„ever"	11	2+8	Almost never + DK	5	Yes (qt: yes/no)
"last year"	1	1+1	Sometimes + Sometimes	5	Yes (qt: yes/no)
"last year"	1	1+2	Sometimes + Almost never	5	Yes (qt: yes/no)
"last year"	1	2+1	Almost never + Sometimes	5	Yes (qt: yes/no)
"last year"	1	2+2	Almost never + Almost never	5	Yes (qt: yes/no)
"last year"	1	3+1	Never + Sometimes	5	Yes (qt: yes/no)

"last year"	1	8+1	DK/NR + Sometimes	5	Yes (qt: yes/no)

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

C_DEMONST_EXTENDED is a dichotomous variable identifying if a source question asks about participation in only demonstration, or also in other activities next to demonstrations, such as marches, protests or sits-in. It takes the value 1 if the source question is extended to other activities next to demonstrations, for example: *During the past three years, did you participate in a protest, march or sit-in?* (ARB/4). If the question asks about protest or demonstrations only, this control variable takes the value 0, for example: *During the last 12 months, have you taken part in a lawful public demonstration?* (ESS/8)

C_DEMONST_ILLEGAL is a dichotomous variable indicating whether the source question wording explicitly refers to the illegal character of the demonstration. It takes the value 1 if respondents were asked about participation in illegal / unlawful / unauthorized demonstrations *During the last 12 months, have you participated in illegal protest activities?* (ESS/1). Otherwise, it takes the value 0.

C_DEMONST_SET indicates whether we used for harmonization a set of source questions from a given survey . It takes the value 1 if we took for harmonization a set of source questions from one survey that specified different time frames. For example: in AMB/2006 one item asked about participation in demonstrations in the last year and the other one asked about having participated in demonstrations ever: PROT1 *Have you ever participated in a public demonstration or protest?* and PROT2 *During last year, did you participate in a public demonstration or protest?* (AMB_2006).

It takes the value 2 if we took a set of source items which asked about different reasons for participation. For example: in ISSP/1996 one question (v10) asks about the protest against the government and another asks about protest, march or demonstration in general (v11).

It takes the value 3 if we took a set of source items which contained questions about legal and illegal demonstrations, for example: in LB/2007 one question asked about participation in authorized demonstrations (p62st.b) and another question asked about participation in protest without authorization (p62n.c).

It takes the value 0 if a single question in a given survey was taken to construct the harmonized variable T_DEMONST.

C_DEMONST_ISSUE is a nominal variable that indicates whether the source question wording deals with specific cause/concern for demonstrating. It takes the value 1 if items asked only about protests related to environmental issues, which happens in six survey waves whose data we harmonized: EB/25, EB/29, EB/58, ISSP/1993, ISSP/2000 and ISSP/2010. It takes the value 2 if source items referred to protesting during the Arab Spring, which is the case of ARB/2. It takes

the value 3 if the source item refers to solving problems in a neighborhood, which is the case of LB/2001.

C_DEMONST_PROTEST is a dichotomous variable identifying whether the term "protest" appears explicitly in the source item wording. It takes the value 1 if the source item contains the term "protest", for example: *Have you ever participated in a public demonstration or protest?* (AMB/2004). Otherwise, C_DEMONST_PROTEST takes the value 0.

3.4. Quality Control Variable for processing error types

QP_DEMONST, stored in the SDR2 PLUG_SURVEY file, measures processing error types that we identified when comparing information about the source variables available in the survey documentation (codebook, questionnaire) with information about the same source variables in data records in the source data files (data dictionaries). Put differently, QP_DEMONST accounts for inconsistencies between different metadata elements, such as data records on the one hand (i.e., variable values, variable labels, value labels - in the source data files), and information in the codebook or questionnaires, on the other hand.

QP_DEMONST is a nominal variable that can take values from 0 (no processing errors) to 13. Values larger than 0 identify what type, or combination of types, of processing errors occur in the source variables associated with T_DEMONST.

Generally, in the SDR2 project we check for five types of processing errors that we define as follows:

- **Illegitimate Variable Values** captures values of the source variable that are outside of the range that SDR proposes as acceptable, e.g., respondents' number of education years exceeding 25, and being 6 years old when they completed full-time education.
- **Misleading Variable Values** shows that variable values, as coded in the questionnaires/codebooks, are not congruent with the data records in the source data files.
- **Contradictory Variable Values** captures inconsistencies in how the same variable value is labeled in a codebook vs. a questionnaire or a data dictionary, such as when "don't know" value is recorded as 88 in the codebook vs. 888888 in the data dictionary.
- **Variable Values Discrepancy** concerns situations when several or all values in the same variable are inconsistently labeled in codebook, vs. questionnaire or a data dictionary (i.e., there are reversed scales or contradictory value labels for more than one variable value), e.g., age of responds when finished education is recorded as year ranges in the questionnaire vs. as years in the data dictionary.

- **Lack of Variable Value Labels** concerns undefined "nulls" and codes (variable values) in the source data file that are not explained in any source of documentation defining variables and their values (i.e., "wild codes: for a given variable, such as value 7 in data records while the documentation describes the variable as a 1 to 5 scale).

Correspondingly, the full list of SDR2 codes for Processing Error Types is:

0 = No processing errors
1 = Illegitimate variable values
2 = Misleading variable values
3 = Contradictory variable values
4 = Variable values discrepancies
5 = Lack of variable value labels
6 = Illegitimate & misleading variable values
7 = Illegitimate & contradictory variable values
8 = Illegitimate & discrepancy variable values
9 = Illegitimate & lack of variable value labels
10 = Contradictory variable values & lack of variable value labels
11 = Discrepant & lack of variable value labels
12 = Illegitimate & contradictory & lack of variable value labels
13 = Illegitimate & discrepant & lack of variable value labels

Values 6 to 13 indicate that a combination of different processing error types occurs within one source variable.

The source variables that we used to create a target `T_DEMONST` had contradictory variable values (`QP_DEMONST` = 3), lack of variable value labels (`QP_DEMONST` = 5) or a combination of these two processing errors (`QP_DEMONST` = 10). Other types of processing errors did not occur in source data used for this target variable, hence the realized values in the `QP_DEMONST` variable are 0, 3, 5 and 10.

4. Special Cases

- CNEP/4 Spain, Indonesia, Mexico, and South Africa data files have two variables concerning participation in demonstration: one corresponding precisely to the question in the national questionnaire, and the other one, its recode. The difference between these two variables concerns the answer options "no" and "don't know." In the original, national variables, "no" and "don't know" are mutually exclusive, while the recoded variables merge them into one category ("0 = no or don't know"). Since merging these two options means losing crucial information, we took the original national variables as source variables, although meta-information concerning them is in national languages.

- ISSP/2004 and 2014 variables contain two answer options: "Have done it in the past year" and "Have done it in the more distant past". We took both options into `T_DEMONST` and created the

control variable concerning time frame for past participation C_DEMONST_YEARS = 1 and C_DEMONST_YEARS = 11 at the individual level. See also Section 3.3.

- In LB/2007, 2008, 2013, and 2015, questionnaire items provide abbreviations instead of meaningful labels. We extrapolate from information earlier LB waves provide (e.g., LB/2006) and assume that HD = "Have done", WD = "Would do", and WND = "Would never do."

- ISSP/1996 contains two questions on protest that we take to construct the target (source variable names: V10 and V11, C_DEMONST_SET = 1). The source item V10 specifies that the protest was "against government." Since protesting government involves a wide range of potential reasons, from tax policies to institutional change or immigration policy, we treat this as a broad cause/concern and code C_DEMONST_ISSUE = 0.

- NEB/1 and NBB/1-3 contain a series of questions that ask respondents how likely they are to participate in demonstrations around certain issues (e.g., NEB/1 "How likely is it that you will take part in such a demonstration? Very likely, Maybe, Not very likely, Definitely not"). These types of question ask only about prospective behavior. In doing so, they are substantively different from the source items that ask respondents about past participation, and about past participation and future willingness. Hence, we decided to exclude these items from the MASTER file.

- WVS contains two variables on participation in demonstrations. One item (source variable name: E027), available in all WVS waves, asks: "Whether you have done any of these things, whether you might do it or would never under any circumstances do it (…): Attending peaceful demonstrations." Since this question does not specify any time boundary for past behavior, the control variable C_DEMONST_YEARS takes the value 11. Waves 5 and 6 provide also a second item, labeled "Political action recently done: Attending peaceful/lawful demonstrations" (source variable name: E221B). In wave 5 this additional question asks about participation during the last 5 years. Hence we code C_DEMONST_YEARS = 5. In wave 6, the second question asks about participation during last year, so C_DEMONST_YEARS = 1 (see also Table 3 for details on recodes for WVS/6; the same procedure applies to WVS/5).

- In WVS/6 the question on participating last year is filtered by the question on participating "ever" (i.e., the second question is asked only to respondents who answered that they did participate in demonstrations). In WVS/5 such a filter is not present.

- In WVS/6 the question about participating in the last year is formulated in terms of frequency of participation, as follows: "how often you have done it [i.e., attending peaceful demonstrations] in the last year" with answer options "not at all," "once," "twice," "three times, "more than three times." We code answers of once or more as *yes*, and T_DEMONST = 5 (question type: *yes/no*). For respondents who answer "not at all" to this question yet answered "yes" to the first question about participation "ever," T_DEMONST = 1 (question type: *yes/no*), cf. rule that the "strongest answer" wins.
- In AFB/1 and WVS there are differences in question wording and response categories for some countries within the same variable, which results in our different target coding schema and control variables.

- In two EB waves (EB/58.2 and EB/62.2) the answer category “do not know” is captured as a separate source variable. To code a target variable, we include the “do not know” source variable. As this case appears only twice in the dataset and is purely technical, the control variable “set of questions” does not distinguish it (see code 0 in the control variable “set of questions”, which means that to construct a target we used a single item and, exceptionally, additional item capturing don't know.

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.