# SDR 2.0
# Control Variables and CWT

# Survey Harmonization ex-post: WORKFLOW

Creating harmonized measures *ex-post* **strongly relies** on released survey data and documentation linked to it.

1. Theoretical model and concept specification
2. Criteria for data selection -> Time zero: data and documentation downloaded
3. Source variables: first (inclusive) selection; item availability template
4. (Re)Define target variable based on source variables; final source variable list
5. Creating harmonization control variables
6. Cross-walk coding and additional source variables check
7. Quality control variables for items, documentation and data
8. Final revisions of harmonization and its documentation; general variable report

# Step 5. Creating Harmonization Control Variables

**<u>Definitions:</u>**

- Control variable (C): a variable controlling specific features of source variable(s) and potentially influencing validity and reliability of the Target variable (T);

- Target variable (T) - variable created through the ex-post harmonization, which is a function of one or more source variables: $T = f(S)$;

- Source variable (S) - variable collected from an original survey for the ex-post harmonization.

In other words: "Harmonization controls $w.j$ are created during the process of converting original variables $s_{ij}$ into the final target variables $T_{ij}$ to capture important features of source items that could introduce varying amounts of measurement error."

**<u>Aims:</u>**

- controlling for the aspects of the intersurvey variability;

- preserving maximum amount of information from the source data in the target variables.

# Target Variables

## Demographics:

- Respondent's Gender
- Respondent's Age
- Birth year of the Respondent

## Political Participation:

- Participation in Demonstrations
- Signing Petitions
- Participation in Boycotts
- Interest in Politics
- Discussing Politics

## Trust:

- Trust in Government
- Trust in the Legal System
- Trust in Parliament
- Trust in Political Parties
- Trust in People

## DWellbeing:

- Subjective Feeling of Happiness
- Subjective Feeling of Life Satisfaction
- Health (general)
- Health (mental)

# Control Variables in Detailed Variable Report



| | B | D | G | J | Q | R | S | T | U | V |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Data file** | **QUESTION WORDING** Priority: 1 = questionnaire 2 = codebook] | **VARIABLE LABEL** Priority: 1 = data dictionnary 2 = codebook | **VARIABLE VALUES** Priority: 1 = data dictionnaries 2 = codebook 3 = questionnaire | C_DE MON ST_YE ARS | C_DE MON ST_EX TEND ED | C_DE MON ST_ILL EGAL infor | C_DE MON ST_SE T: | C_PR_ DEM ONST _ISSU E 1- | C_D MO ST_F OTE 1=th |
| 11 | ABS_4_SG | Here is a list of actions that people sometimes | *76a Attended a demonstration or* | -1\|Missing | 3 | 0 | 0 | 0 | 0 | 1 |
| 12 | ABS_4_TH | Here is a list of actions that people sometimes | *76a Attended a demonstration or* | -1\|Missing | 3 | 0 | 0 | 0 | 0 | 1 |
| 13 | ABS_4_TW | Here is a list of actions that people sometimes | *76a Attended a demonstration or* | -1\|Missing | 3 | 0 | 0 | 0 | 0 | 1 |
| 14 | AFB_1 | Question text - SAB | pardemo/Attend demonstration | 0\|Never | SAB:11 | 0 | 0 | 0 | S0 | 1 |
| 15 | AFB_2 | Here is a list of actions that people sometimes | Q25d. Attend a demonstration or | -1\|Missing | 1 | 0 | 0 | 0 | 0 | 1 |
| 16 | AFB_3 | Here is a list of actions that people sometimes | Q31c. Attend a demonstration or | -1\|Missing | 1 | 0 | 0 | 0 | 0 | 1 |
| 17 | AFB_4 | Here is a list of actions that people sometimes | Q23c. Attend a demonstration or | -1\|Missing | 1 | 0 | 0 | 0 | 0 | 1 |
| 18 | AFB_5 | Here is a list of actions that people sometimes | Q26d. Attend a demonstration or | -1\|Missing | 1 | 0 | 0 | 0 | 0 | 1 |
| 19 | AFB_6 | Here is a list of actions that people sometimes | Q27e. Attend a demonstration or | -1\|Missing | 1 | 0 | 0 | 0 | 0 | 1 |
| 20 | AMB_2004 | Have you ever participated in a public | Have you participated in a | 1\|Algunas veces | 11 | 0 | 0 | 0 | 0 | 1 |
| 21 | AMB_2006 | Have you ever participated in a public | Have you ever participated in a | 1\|Algunas veces | 11 | 0 | 0 | 1 | 0 | 1 |
| 22 | AMB_2006 | During last year, did you participate in a | In the last year, have you | 1\|Algunas veces | 1 | 0 | 0 | 1 | 0 | 1 |
| 23 | AMB_2008 | Have you ever participated in a public | Have you ever participated in a | 1\|Sometimes | 11 | 0 | 0 | 1 | 0 | 1 |

Change log | Dictionary | T_DEMONST | Left-overs

# Types of Control Variables

**Question Wording**
- **NONSTANDARD**
- **EXTENDED**
- **ISSUE**
- **QUESTION SPECIFIC CONTROL VARIABLES**

**Question Context**
- **SET**
- **LEVEL**

**Response Categories**
- **TIME SPAN**
- **SCALE LENGTH**
- **ASCEND**
- **POLARITY**
- **QUESTION SPECIFIC CONTROL VARIABLES**

# Types of Control Variables: Question Wording

- **NONSTANDARD**:

0 = standard question (each variable has a distinct standard question);

n = the question in the source documentation deviates from the standard question;

- **EXTENDED**: dummy variable that indicates whether the source question asked about activities additional to the main actions;

- **ISSUE**: indicates that source items specify the cause or concern for which the action took place.

# Types of Control Variables: Survey Questions. Examples

| | | |
|---|---|---|
| Source non-standard question wording | **NONSTANDARD** | 0 = standard question wording (general interest in politics)<br><br>1 = interest in national / public affairs<br><br>2 = interest in politics and affairs of the nation<br><br>3 = interest in politics and government<br><br>4 = interest in politics and economics<br><br>….<br><br>9 = no source question wording available |
| Source variable question asks about activities additional to the main activation | **EXTENDED** | 0 = otherwise (e.g. participation in demonstrations)<br>1 = item extended (e.g. participation in marches, protests or sits-in) |
| Source item specifies cause/concern for action | **ISSUE** | 0 = issue not specified<br><br>1 = signing petition of specific issue (e.g. environmental problems)<br><br>….<br><br>3 = other |

# Types of Control Variables: Question Wording

| | | |
|---|---|---|
| Source non-standard question wording | **NONSTANDARD** | 0 = standard question wording<br>1 …. n = nonstandard question wording (specified) |
| Source variable question asks about activities additional to the main activation | **EXTENDED** | 0 = otherwise<br>1 = item extended |
| Source item specifies cause/concern for action | **ISSUE** | 0 = issue not specified<br>1 … n = issue specified |

# Types of Control Variables: QUESTION SPECIFIC CONTROL VARIABLES

- **QUESTION SPECIFIC CONTROL VARIABLES**: Control variables designed to control for a special word or issue in the survey question.

E.g.

DEMONST_ILLEGAL: participation in illegal demonstration(s);

TR_LEG_COURTS: question asks about trust in courts;

BOYCOTT_PRODUCTS: participation in boycotts; boycotting certain products.

# Types of Control Variables: QUESTION SPECIFIC CONTROL VARIABLES

| | | |
|---|---|---|
| Source item asks about participation in illegal demonstration | DEMONST_ILLEGAL | 0 = otherwise<br>1= illegal demonstration |
| Question about trust in legal system mentions courts, court system, judiciary | TR_LEG_COURTS | 0 = false<br>1 = true |
| Source item wording explicitly mentions 'boycotted certain products' | BOYCOTT_PRODUCTS | 0 = otherwise<br>1 = boycotted certain products |

# Types of Control Variables: Question Context

- **SET**: two or more source items used to construct the target variable

0 = single item in the questionnaire

1 ... n = set of items (topic specific)

- **LEVEL**: question selected from a set of questions or a single question item

0 = level not specified

1 = selected from a set of questions

2 = only one question available

# Types of Control Variables: Question Context. Examples

| | |
|---|---|
| Two or more source items **SET** used to construct the target variable | 0 = single item<br><br>1 = set of items on legal or illegal participation in demonstrations;<br><br>2 = set of questions on participation in demonstrations for different reasons for participation; |
| Question on interest in **LEVEL** politics in general or on the national level | 1 = national, selected from the set<br><br>2 = national, only one question available<br><br>0 = level not specified (general interest in politics) |

# Types of Control Variables: Response Categories.

| | | |
|---|---|---|
| Source variable action time span | **TIME SPAN** | 1 = last year<br>2 = 2 years<br>...<br>n = n years<br>11 = ever |
| Source variable scale length | **SCALE LENGTH** | 2 = 2-point scale<br>...<br>11 = 11-point scale |
| Source never discussing politics with | **POLITICS_NEVER_WITH** | 1 = friends<br>2 = friends or family/relatives<br>...<br>0 = unspecified |
| Source variable scale direction | **ASCEND** | 1 = ascending -> responses are **ordered** from 'lowest' to 'highest'<br>0 = otherwise, i.e. descending |
| Source variable - polarity | **POLARITY** | 1 = unipolar (from min to max of one item: min trust->max trust)<br>2 = bipolar (from min of one item to max of different item: min distrust->max trust)<br>3 = bipolar (agree-disagree) |

# Survey Harmonization ex-post: WORKFLOW

Creating harmonized measures *ex-post* **strongly relies** on released survey data and documentation linked to it.

1. Theoretical model and concept specification

2. Criteria for data selection -> Time zero: data and documentation downloaded

3. Source variables: first (inclusive) selection; item availability template

4. (Re)Define target variable based on source variables; final source variable list

5. Creating harmonization control variables

6. Cross-walk coding and additional source variables check

7. Quality control variables for items, documentation and data

8. Final revisions of harmonization and its documentation; general variable report

# Step 6. Cross-walk coding and additional source variables check

**AIM: Maping Source Variables to Target Variables**

- CWTs or Cross-Walk Tables are macro-enabled Excel documents, the basis for harmonization syntax
- ➤ Contain detailed information on recodes from source to target values;
- ➤ Serve for additional consistency checks and enables quick insight into data (national survey level distributions for each source variable).
- CWT parts:
- ➤ Names of target variables;
- ➤ Source variable values;
- ➤ Distribution of value by country;
- ➤ Values of target variables;

# SDR 2.0 Format: CWT - The Cross-Walk Table

C2842

| | A filter variable | | | | | | select | | B item # | | C target label | | prepare | make all SQL | | D parameter 1 | | check src value | clear | | E parameter 2 | | make SQL | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | « | ‹ | › | » | | | ▼ | | ▼ | | | | | ▼ | | | | | ▼ | | | | ▼ |
| 77 | ABS_4_KR q76a | | | | | | | | 1 | | | | | | | **ABS_4_KR q76a** | | | | | | | | |
| 78 | ABS_4_KR q76a | | | | | | | | 2 | | | | | | | **76a Attended a demonstration or protest march-4 categories** | | | | | | | | |
| 79 | ABS_4_KR q76a | | | | | | | | 3 | | miss | | | | | -1 | | | | | Missing | | | |
| 80 | ABS_4_KR q76a | | | | | | | | 3 | | yes (yes\|would\|would not) | | | | | 1 | | | | | I have done this more than once | | | |
| 81 | ABS_4_KR q76a | | | | | | | | 3 | | yes (yes\|would\|would not) | | | | | 2 | | | | | I have done this once | | | |
| 82 | ABS_4_KR q76a | | | | | | | | 3 | | would (yes\|would\|would not) | | | | | 3 | | | | | I have not done this, but I might d | | | |
| 83 | ABS_4_KR q76a | | | | | | | | 3 | | would not (yes\|would\|would not) | | | | | 4 | | | | | I have not done this and I would n | | | |
| 84 | ABS_4_KR q76a | | | | | | | | 3 | | miss | | | | | 7 | | | | | Do not understand the question | | | |
| 85 | ABS_4_KR q76a | | | | | | | | 3 | | dk | | | | | 8 | | | | | Can't choose | | | |
| 86 | ABS_4_KR q76a | | | | | | | | 3 | | ref | | | | | 9 | | | | | Decline to answer | | | |
| 87 | ABS_4_KR q76a | | | | | | | | 4 | | | | | | | ABS_4_KR | | | | | q76a | | | |
| 88 | ABS_4_KR q76a | | | | | | | | 5 | | | | | | | 1 2 3 4 8 9 | | | | | case when q76a = '1' then 4 whe | | | |
| 89 | ABS_4_MM q76a | | | | | | | | 1 | | | | | | | **ABS_4_MM q76a** | | | | | | | | |
| 90 | ABS_4_MM q76a | | | | | | | | 2 | | | | | | | **76a Attended a demonstration or protest march-4 categories** | | | | | | | | |
| 91 | ABS_4_MM q76a | | | | | | | | 3 | | miss | | | | | -1 | | | | | Missing | | | |
| 92 | ABS_4_MM q76a | | | | | | | | 3 | | yes (yes\|would\|would not) | | | | | 1 | | | | | I have done this more than once | | | |
| 93 | ABS_4_MM q76a | | | | | | | | 3 | | yes (yes\|would\|would not) | | | | | 2 | | | | | I have done this once | | | |
| 94 | ABS_4_MM q76a | | | | | | | | 3 | | would (yes\|would\|would not) | | | | | 3 | | | | | I have not done this, but I might d | | | |
| 95 | ABS_4_MM q76a | | | | | | | | 3 | | would not (yes\|would\|would not) | | | | | 4 | | | | | I have not done this and I would | | | |
| 96 | ABS_4_MM q76a | | | | | | | | 3 | | miss | | | | | 7 | | | | | Do not understand the question | | | |

memo | dictionary | **cross-walk** | ⊕

# SDR 1.0 Format: DVR



| E9 | | | $f_x$ | 1 Some times | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|

| | A | B | C | D | E | F | H | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Data file | Source variable name | Item number in questionnaire | Questionnaire item | Source variable values **S_PR_DEMONST** | Fact | t_pr_demonst_fact: Source values recoded as 1 (true) | C_PR_DEMONST_YEARS | C_PR_DEMONST_EXTENDED note: *country specific | C_PR_DEMONST_ILLEGAL note: *country specific categories | C_PR_DEMONST_SET_OF_QUESTIONS note: |
| 5 | AFB_1 | pardem | 0 | Question text - SAB | 0 Never, 1 Only | 1 | 1 2 3 | 5 | 1* | 0* | 0* |
| 6 | AFB_2 | q25d | Q25D | Here is a list of | 0 No, would never | 1 | 2 3 4 | 1 | 1 | 0 | 0 |
| 7 | AFB_3 | q31c | q31c | Here is a list of | 0 No, would never | 1 | 2 3 4 | 1 | 1 | 0 | 0 |
| 8 | AFB_4 | Q23C | Q23C | Here is a list of | 0 No, would never | 1 | 2 3 4 | 1 | 1 | 0 | 0 |
| 9 | AMB_2004-2012 | prot1 | 0 | 2004- Have you | 1 Some times | 1 | 1 2 | | 1* | 0* | 0* |
| 10 | AMB_2004-2012 | prot3 | 0 | 2010 In the last 12 | 1 Yes  2 No | 1 | 1 | 1 | 1 | 0* | 0* |
| 11 | ARB_1 | q2302 | q2302 | Here is a list of | 1 Once | 1 | 1 2 | 3 | 1 | 0 | 0 |
| 12 | ARB_2 | q5022 | Q502.2 | Here is a set of | 0 0. missing | 1 | 1 2 | 3 | 1* | 0* | 0* |
| 13 | ASES | V0170 | Q405f | Now I'd like you to | 1 Have often done | 1 | 1 2 | | 1* | 0* | 0* |
| 14 | CNEP_3_ES | Z.Sp.H.Protest | P. 70 | P.70. Have you | 1 Yes | 1 | 1 | 1 | 0 | 0 | 0 |
| 15 | CNEP_3_MX | Z.Me.H.ProtestPre | 41 | 41. HAVE YOU | 1 Yes | 1 | 1 | | 0 | 0 | 0 |
| 16 | CNEP_3_MZ | H.Protest | 47 | Q. 47A Did you | 0 No | 1 | 1 | 1 | 0 | 0 | 0 |
| 17 | CNEP_3_PT | H.Protest | 65 | P 65– Have you | 1 Yes | 1 | 1 | | 1 | 0 | 0 |
| 18 | CNEP_3_TW | H.Protest | N17 | N17. Have you | 1 Yes | 1 | 1 | 1 | 0 | 0 | 0 |

INFO    **SOURCE VARIABLES**    AVAILABILITY

# Conclusions

| For the Data Users | For the Data Harmonizers |
|---|---|
| • Differences in question wording;<br>• Differences in question context;<br>• Differences in types of scales;<br>• Differences in scale direction. | • Creation of the detailed and structured documents to store the information from the source data;<br>• Preserving maximum amount of information;<br>• Separate internal documents;<br>• Clear structure of internal documents;<br>• Preservation of internal comments;<br>• Clear concepts;<br>• Attention towards question wording and cafeteria of answers;<br>• Creation of user friendly output. |