

International Training Event on Survey Methodology

Cross-national Survey Harmonization and Analysis: Weights, Data Quality and Multi-level Modeling

May 11-16, 2015, OSU Main Campus (Columbus, OH)

Session I: Introduction

**Cross-national Studies: Interdisciplinary Research and Training Program - CONSIRT,
consirt.osu.edu**

the Polish Academy of Sciences (PAN) and The Ohio State University (OSU)

Mershon Center for International Security Studies

OSU Departments of Sociology and Political Science

OSU Polish Study Initiative

Presenters:

Weights – Introduction by Kazimierz M. Slomczynski

- **Marcin Zieliński, Polish Academy of Sciences, University of Warsaw**

Data Quality Indicators - Introduction by Kazimierz M. Slomczynski

- **Matt Schoene, Ohio State University**
- **Marta Kołczyńska, Ohio State University**
- **Ilona Wysmułek, Graduate School for Social Research (Warsaw)**
- **Olena Oleksiyenko, Graduate School for Social Research (Warsaw)**
- **Przemek Powałko, Polish Academy of Sciences**

Multi-level modeling - Introduction by Kazimierz M. Slomczynski

- **Robert Kunovich, University of Texas at Arlington**

Multi-survey data analysis

Goal of the training event: Bringing weights for units of observation and data-quality indicators into analysis by means of multi-level modeling

Multi-survey data structure:

(1) Survey project (SP): a collection of national surveys under a common administrative umbrella (e.g., European Social Survey, World Values Survey)

Differentiation of SPs with respect to thematic focus and organizational features

(2) Within SPs are project waves (PW), which contain the surveys conducted at the same time according to the same instrument

PWs differ within SPs

(3) Within PW are national surveys (NS) conducted in given countries or “national territories”

NSs differ within PWs

(4) NSs include individuals (ID)

Weights and data-quality indicators

- **In the context of multi-survey data, weights are numerical values of the importance of a given unit of observation: an individual (ID), national survey (NS), wave (PW), project (SP).**
- **Data-quality indicators can deal with records for individuals (ID), documentation for each survey (NS), documentation of data for a given wave (PW) and properties of a survey project (SP).**

LEVEL	WEIGHTS	DATA QUALITY
SURVEY PROJECT (SP)	Analytic and importance weights: selection and clustering	Quality of cross-national coordination and control
WAVE (PW)		Quality of data description
NATIONAL SURVEY (NS)	Sample-size weights Population-size weights	Quality of survey administration
INDIVIDUAL (ID)	Design weights Post-stratification weights	Quality of computer records

Using weights and data-quality indicators

Explanation of attitudes and behaviors by variables on individual level and higher-order level(s)

E.g.: DV = trust in parliament (Y_{ijk}) by individuals i in surveys j from project-waves k

IVs: gender (X1), age (X2), education (X3), type of political regime (measured by political rights, Z1), and economic development (GDP per capita, Z2)

- $Y_{ijk} = f(X_{ijk}, Z_{.jk})$

Including Weights (W) and Data-Quality Indicators (Q)

- $Y_{ijk} = f(X_{ijk}, Z_{.jk}, W_{ijk}, W_{.jk}, W_{..k}, Q_{ijk}, Q_{.jk}, Q_{..k})$

Data Recycling of International Survey Projects

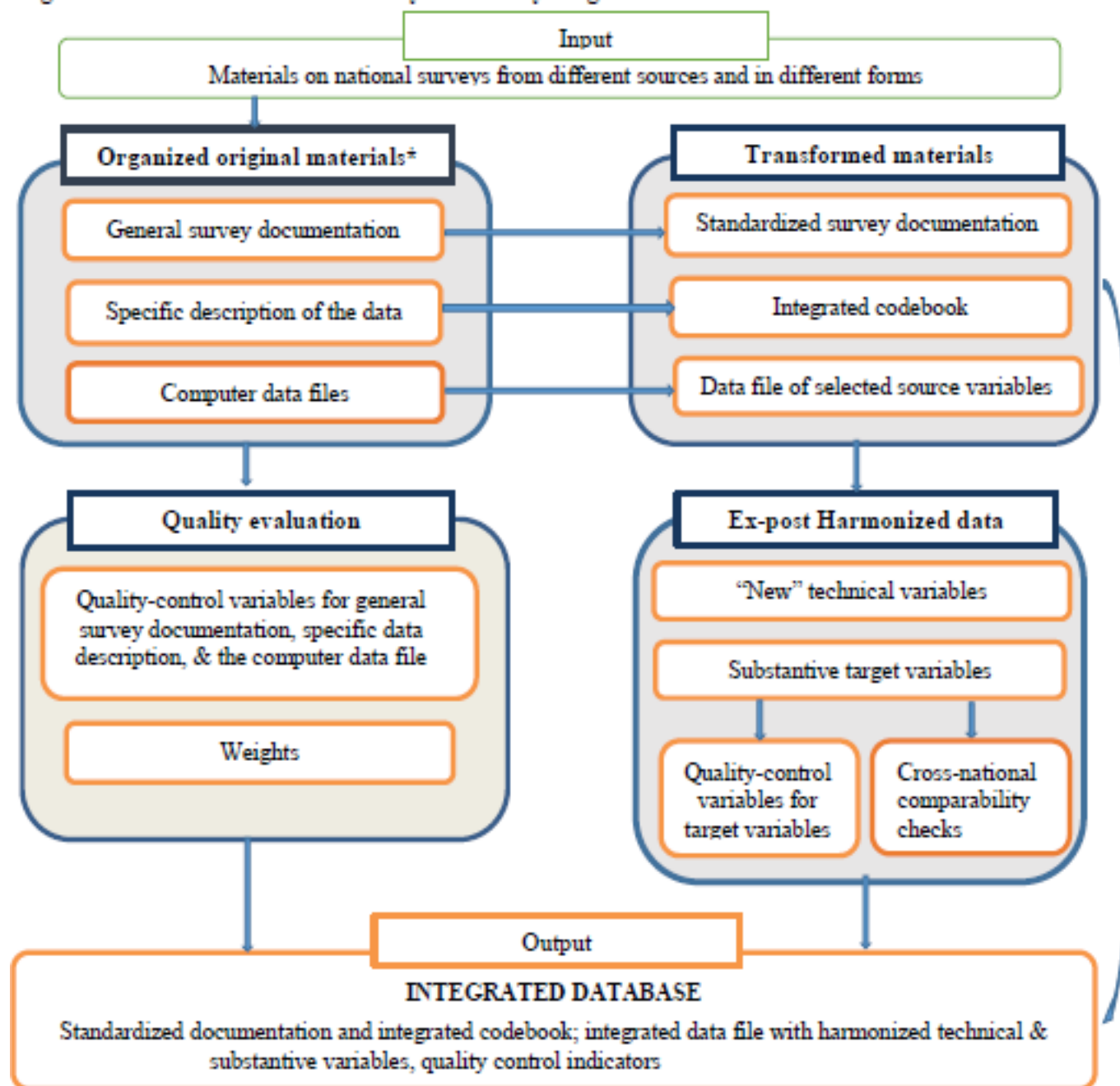
Survey Data Recycling (SDR)

- SDR = framework for (re)-processing cross-national survey data;
- SDR concerns survey data quality control & *ex-post* survey harmonization, to:
 - (a) account for “messiness” of the original source data,
 - (b) provide “comparable” data

Thus, SDR expands the scope of extant projects (time, space, number of observations, types of indicators)

Unifying thesis: account for errors & biases in original surveys & harmonization procedures *via* different types of quality control variables, to be included in substantive analyses.

Figure 1. General Schema of Survey Data Recycling



Criteria of selecting survey projects

- containing questions about political attitudes and behaviors**
- designed as cross-national, and, preferably, multi-wave;**
- with the samples intended as representative of the adult population of given country or territory;**
- non-commercial;**
- freely available in the public domain;**
- with documentation - study description, codebook and/or questionnaire - in English**

Abbrev.	Survey Project	Time span	Waves	Files	Data Sets	Cases
			Counts			
AFB	Afrobarometer	1999-2009	4	4	66	98942
<u>AMB 1</u>	Americas Barometer	2004-2012	5	1	92	151341
ARB	Arab Barometer	2006-2011	2	2	16	19684
ASB	Asian Barometer	2001-2011	3	3	30	43691
<u>ASES 2</u>	Asia Europe Survey	2000	1	1	18	18253
<u>CB 3</u>	Caucasus Barometer	2009-2012	4	4	12	24621
<u>CDCEE 4</u>	Consolidation of Democracy	1990-2001	2	1	27	28926
<u>CNEP 5</u>	Comparative National Elections Project	2004-2006	1	8	8	13372
<u>EB 6</u>	Eurobarometer	1983-2012	7	7	152	138753
<u>EQLS 7</u>	European Quality of Life Survey	2003-2012	3	1	93	105527
<u>ESS 8</u>	European Social Survey	2002-2013	6	2	146	281496
<u>EVS_9</u>	European Values Study	1981-2009	9	1	312	423084
<u>ISJP 10</u>	International Social Justice Project	1991-1996	2	1	21	25805
<u>ISSP 11</u>	International Social Survey Programme	1985-2013	13	13	363	493243
<u>LB 12</u>	Latinobarometro	1995-2010	15	15	260	294965
<u>LITS 13</u>	Life in Transition Survey	2006-2010	2	2	64	67866
<u>NBB 14</u>	New Baltic Barometer	1993-2004	6	1	18	21601
PA2	Political Action II	1979-1981	1	1	3	4057
PA8NS	Political Action – 8 Nation Study	1973-1976	1	1	8	12588
PPE7N	Political Participation in 7 Nations	1966-1971	1	7	7	16522
<u>VPCPCE_15</u>	Values/Political Change	1993	1	5	5	4723
<u>WVS 16</u>	World Values Survey	1981-2009	9	1	312	423084

Data

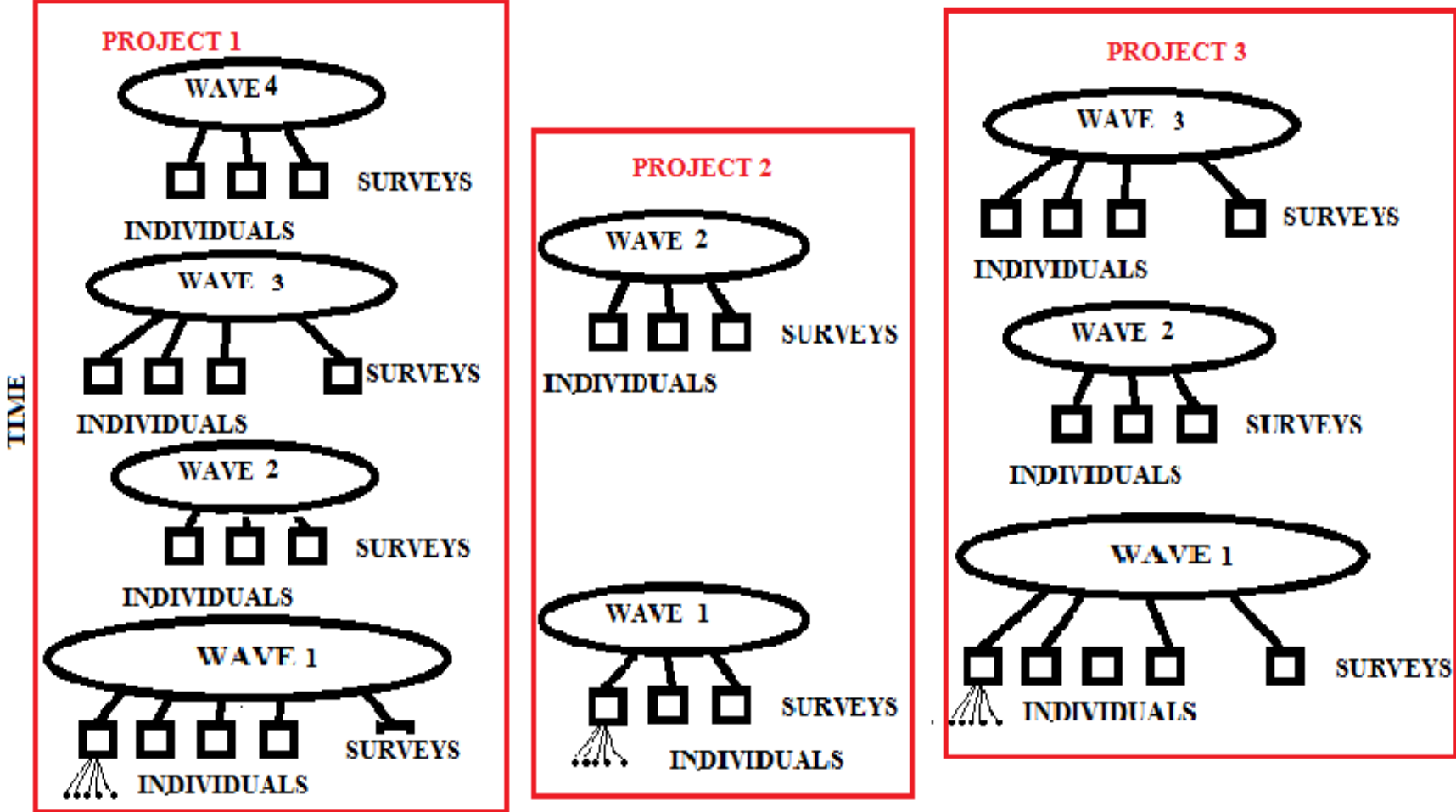
- **1,721 national surveys with over 95% of samples from 500 to 3,000 respondents**
- **National surveys conducted in 142 countries and territories over a period of almost 50 years**
- **All these surveys contain over 25,000 variables**
- **From 150 to 200 variables are identical or similar in large majority of 89 waves**
- **From 25 to 40 variables deal directly with political attitudes and behaviors**

Importance of the SDR project for social sciences: opportunities to study various aspects of attitudes and behaviors

Data for training event

- **European countries + USA**
- **Time: 1989-2013**
- **16 survey projects (SP), each containing – in our dataset – from 1 to 12 waves, altogether 61 project waves (PW)**
- **890 national surveys (NS)**

Data for training event



LEVEL	Number	Comments
SURVEY PROJECT (SP)	16	Different access to data
WAVE (PW)	61	For some projects only one wave; from 1 to 12 per project
NATIONAL SURVEY (NS)	890	NSs contain surveys conducted in national territories; average sample = 1,368 cases
INDIVIDUAL (ID)	1,217,455	-

Formal framework for studying the effect of weights

1. Relationship between target variables T and source variables S

$$T = f(S) \quad [\text{substantive decision of } f]$$

2. Relationship between T and X and weights W

X = Substantive independent variables

W = Weights

$$(1) T = b_0 + b_1 X + e \quad \text{without weights}$$

$$(2) T = b_0 + b_1 X + e \quad \text{with weights}$$

If e is negligible & b_1 in (1) = b_1 in (2) no effect of weights.

The impact of weights on standard errors.