



# Sampling Schemes and Survey Quality in Cross-national Surveys

**Marta Kołczyńska**

Institute of Philosophy and Sociology, Polish Academy of Sciences  
Cross-National Studies: Interdisciplinary Research and Training Program (CONSIRT)

March 27, 2018

Comparative Survey Design and Implementation Workshop

Mary Immaculate College, Limerick, Ireland

# Purpose

Present a classification of sampling schemes in cross-national surveys.

Apply this classification to 22 cross-national survey projects (1721 national surveys).

Describe the variation in sampling schemes in these projects.

Analyze associations between types of sampling scheme and survey quality indicators.

# Sampling scheme classification

- Non-probability samples
  - Quota samples
  - Random-route samples
- Probability samples
  - Multi-stage samples
    - Address register
    - Individual register
  - Simple / stratified random sample
- No / insufficient information
  - Insufficient information
  - No information

Adopted from Ulrich Kohler's 2007 paper on unit non-response bias.

# Data

Survey Data Recycling (SDR) database, version 1.0  
(available at Harvard's Dataverse)

22 cross-national survey projects meeting the following criteria:

- (1) contain measures of *political attitudes and behaviors*, and main individual-level correlates;
- (2) are non-commercial;
- (3) designed as cross-national, and, preferably, multi-wave;
- (4) national samples are **intended** as representative of the adult population;
- (5) documentation (study description, codebook, questionnaire) is provided in English;
- (6) projects are freely available in the public domain.

<b>Survey Projects</b>	<b>Time span</b>	<b># Waves</b>	<b># Files</b>	<b># Data Sets</b>	<b># Cases</b>
<b>Asian Barometer</b>	2001-2011	3	3	30	43691
<b>Afrobarometer</b>	1999-2009	4	4	66	98942
<b>Americas Barometer</b>	2004-2012	5	1	92	151341
<b>Arab Barometer</b>	2006-2011	2	2	16	19684
<b>Asia Europe Survey</b>	2000	1	1	18	18253
<b>Caucasus Barometer</b>	2009-2012	4	4	12	24621
<b>Consolidation of Democracy (C./East Europe)</b>	1990-2001	2	1	27	28926
<b>Comparative National Elections Project</b>	2004-2006	1	8	8	13372
<b>Eurobarometer</b>	1983-2012	7	7	152	138753
<b>European Quality of Life Survey</b>	2003-2012	3	1	93	105527
<b>European Social Survey</b>	2002-2013	6	2	146	281496
<b>European Values Study/World Values Survey</b>	1981-2009	9	1	312	423084
<b>International Social Justice Project</b>	1991-1996	2	1	21	25805
<b>International Social Survey Programme</b>	1985-2013	13	13	363	493243
<b>Latinobarometro</b>	1995-2010	15	15	260	294965
<b>Life in Transition Survey</b>	2006-2010	2	2	64	67866
<b>New Baltic Barometer</b>	1993-2004	6	1	18	21601
<b>Political Action II</b>	1979-1981	1	1	3	4057
<b>Political Action – 8 Nation Study</b>	1973-1976	1	1	8	12588
<b>Political Participation and Equality in 7 Nations</b>	1966-1971	1	7	7	16522
<b>Values/Political Change, Post-communist E</b>	1993	1	5	5	4723
	<b>1966-2013</b>			<b>1721</b>	<b>2,290,060</b>

# Method

Examine available descriptive survey documentation (method reports, technical reports, study descriptions) of 1721 surveys in the SDR dataset to identify information on sampling design.

Pull all information on sampling design into a single text file, together with the name of the source document.

Use automated and manual coding to identify sampling types, according to the classification.

# Coding of sampling schemes

- Quota - if quotas were used at any stage, even only within the household
- Random-route
- Multistage probability samples drawn from an address register
- Multistage probability samples drawn from an individual register
- Single-stage samples (incl. simple and stratified random samples)
- Insufficient information: sampling descriptions do not provide enough information to allow coding sampling type as 1 through 5
- No information: survey documentation lacks any sampling description

# Challenges

Projects, and surveys within projects, vary greatly with regard to the the amount of information they provide about the survey process, including sampling.

Example:

ISSP/2011/Slovenia provides a detailed definition of the target population and the sampling method, including the name of the sampling frame.

ISSP/2011/USA: “Multi-stage area probability sample”.



# Results

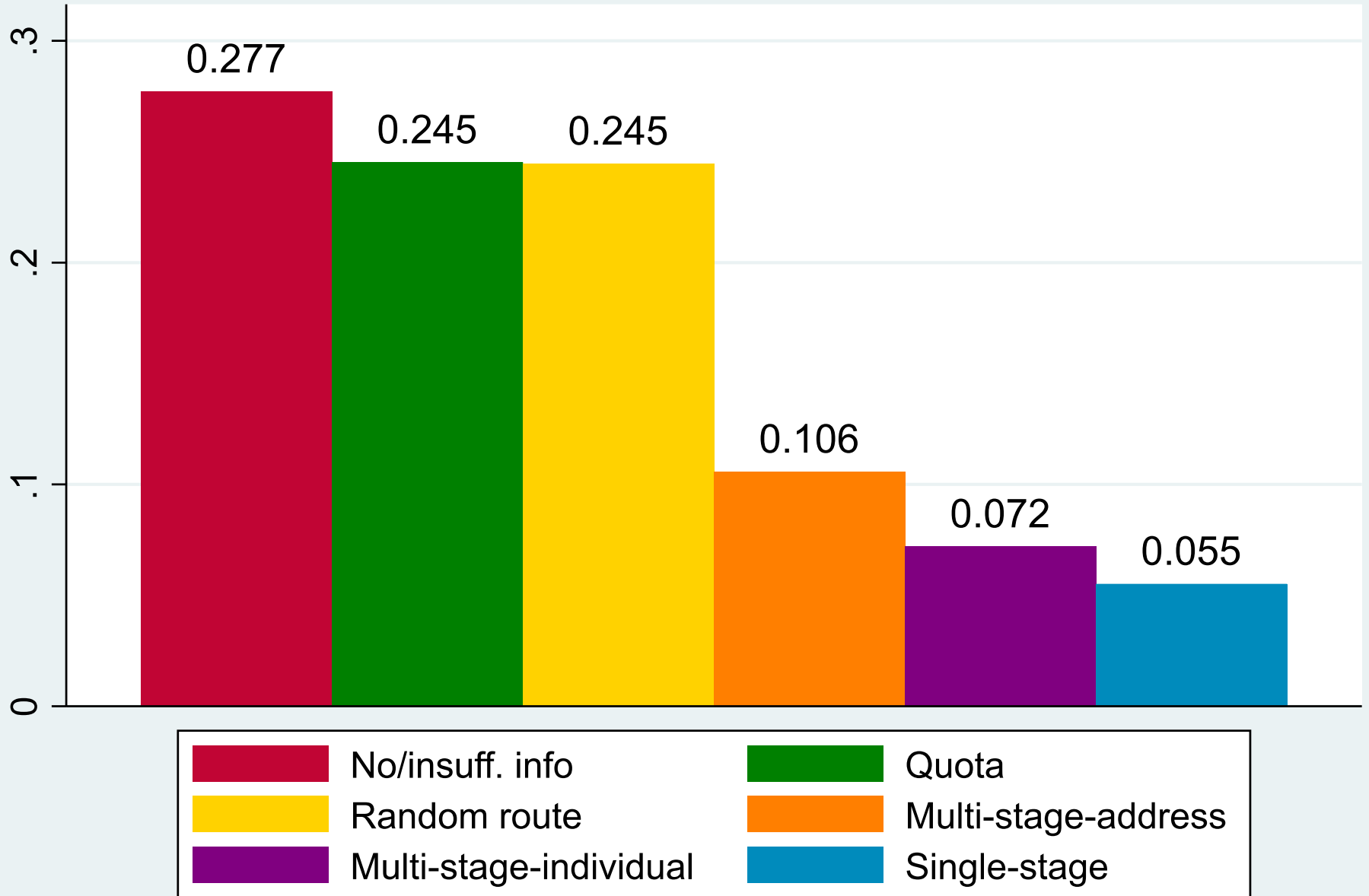
Sampling schemes were coded for each of the 1721 national surveys (project-wave-country).

Variation in three dimensions:

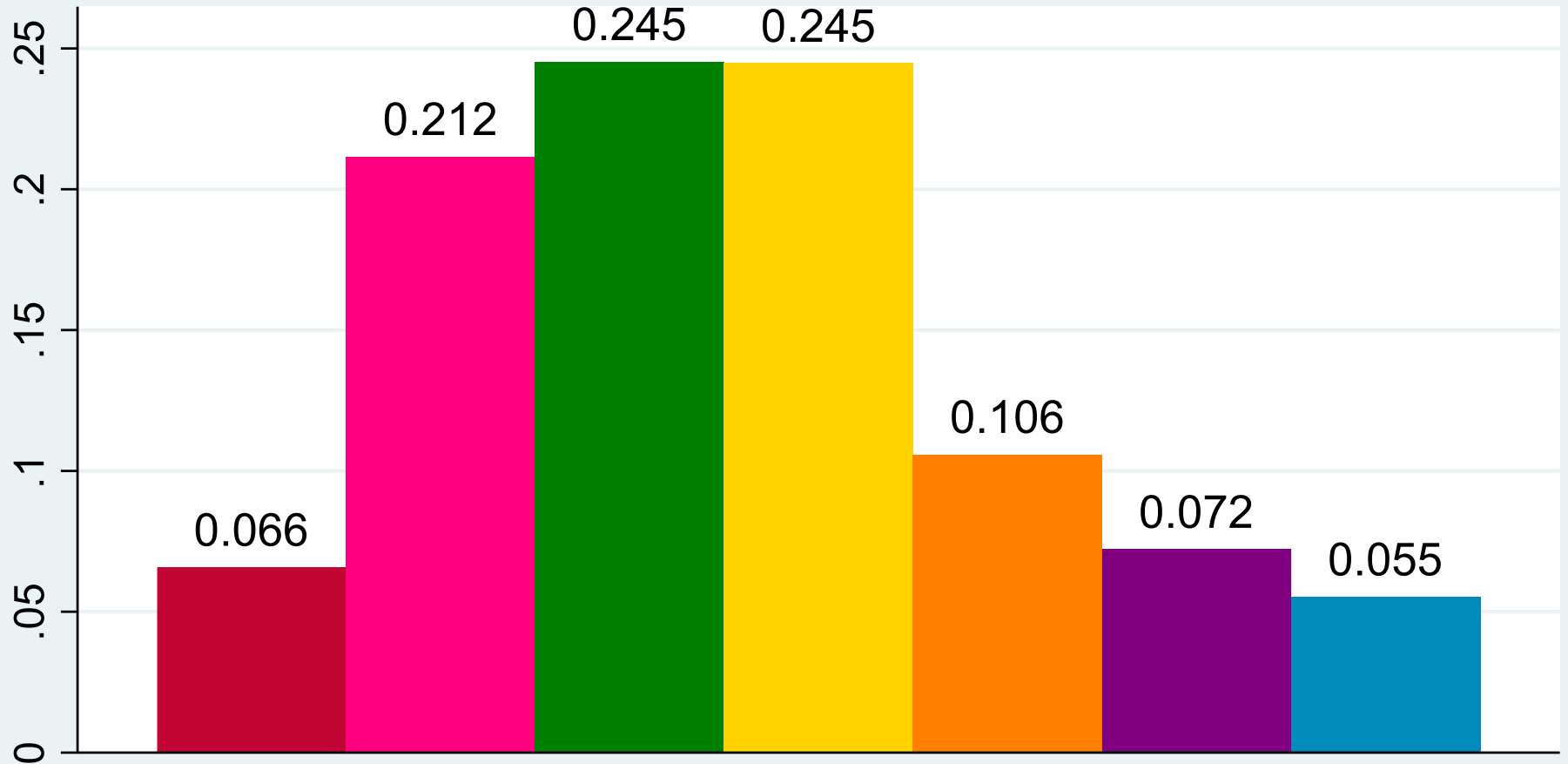
- project
- space
- time

corresponding to three ways of explaining the variation in sampling schemes.

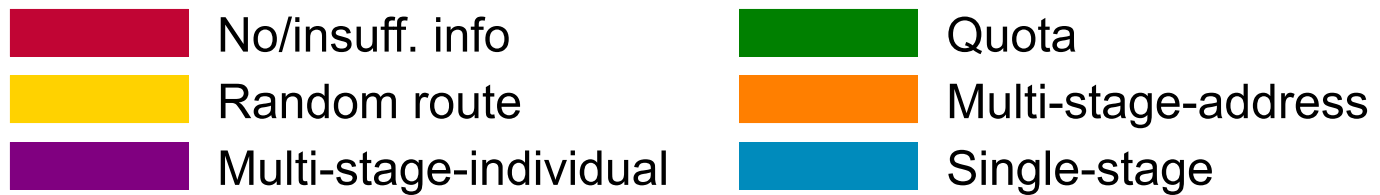
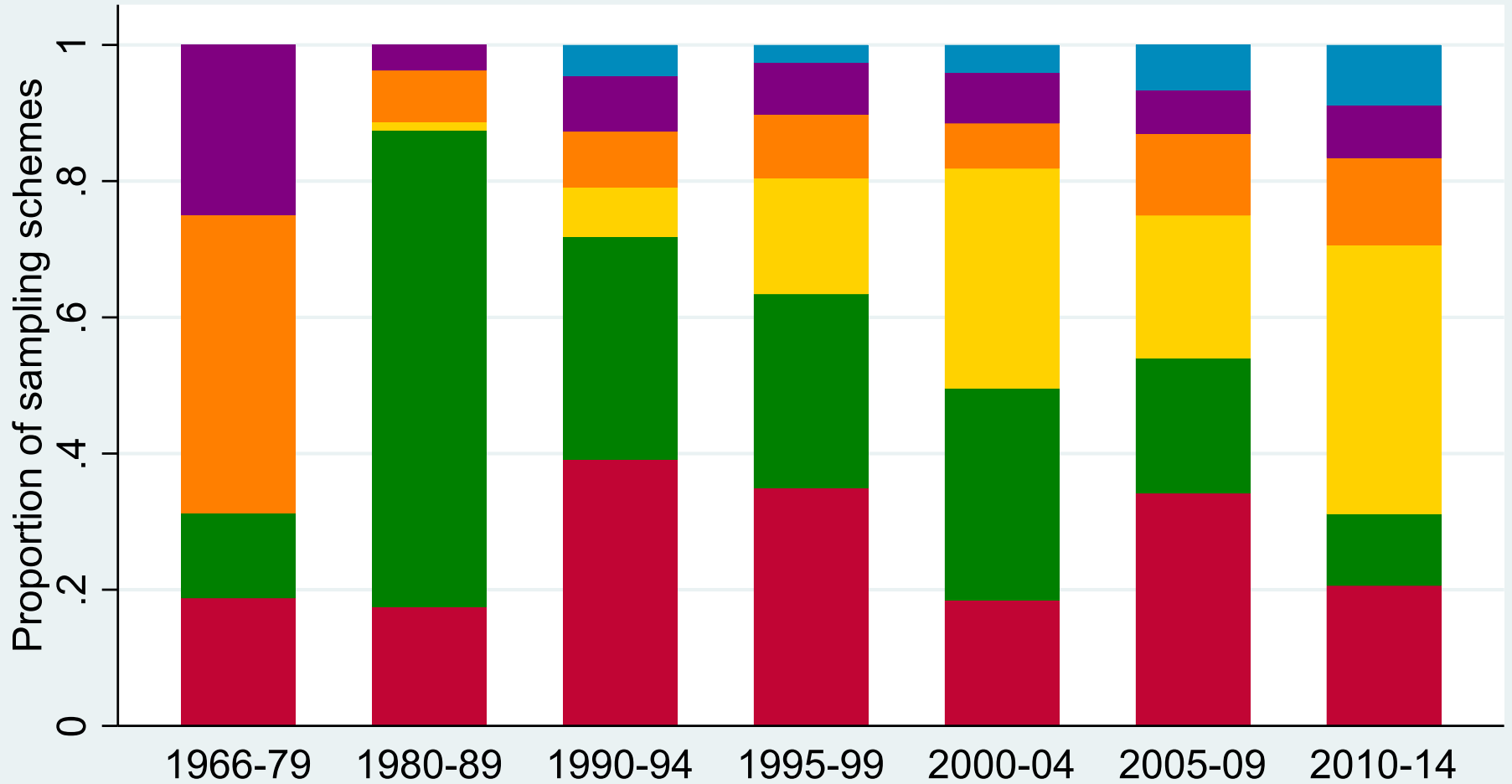
## Overall distribution (6 categories)



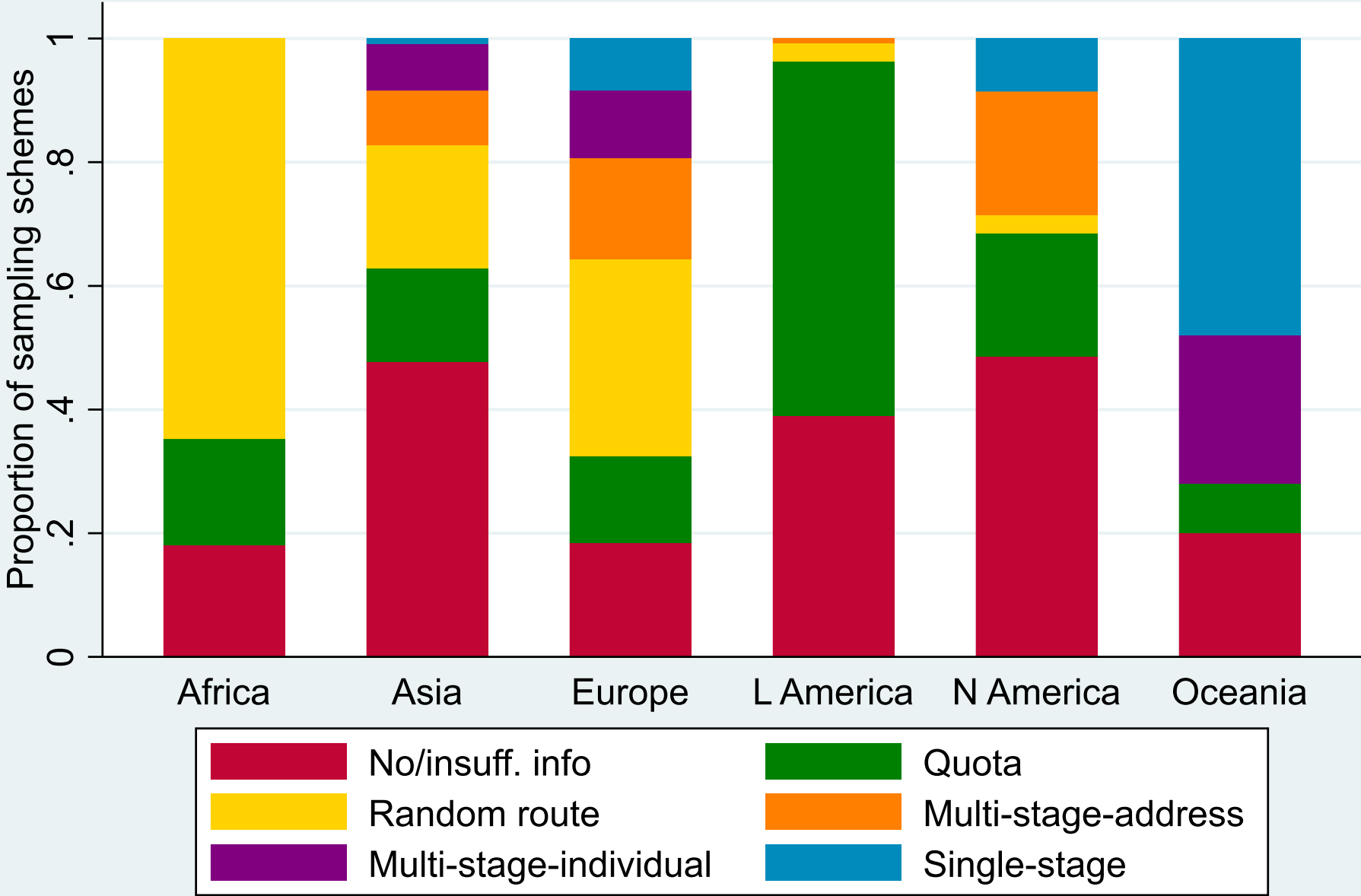
## Overall distribution (7 categories)



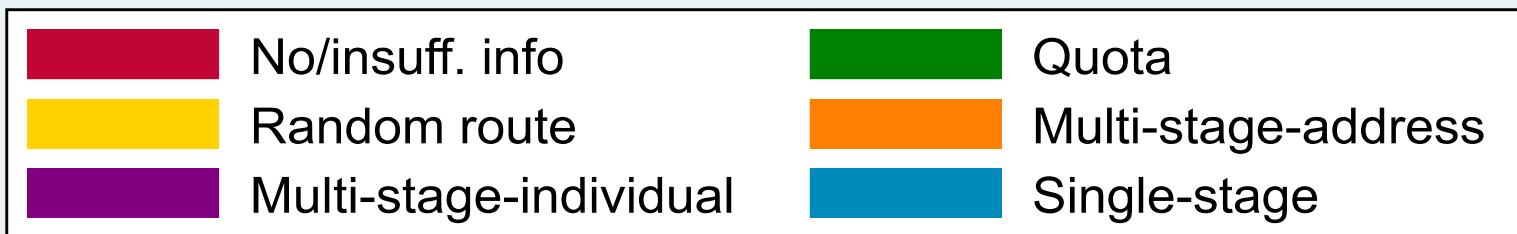
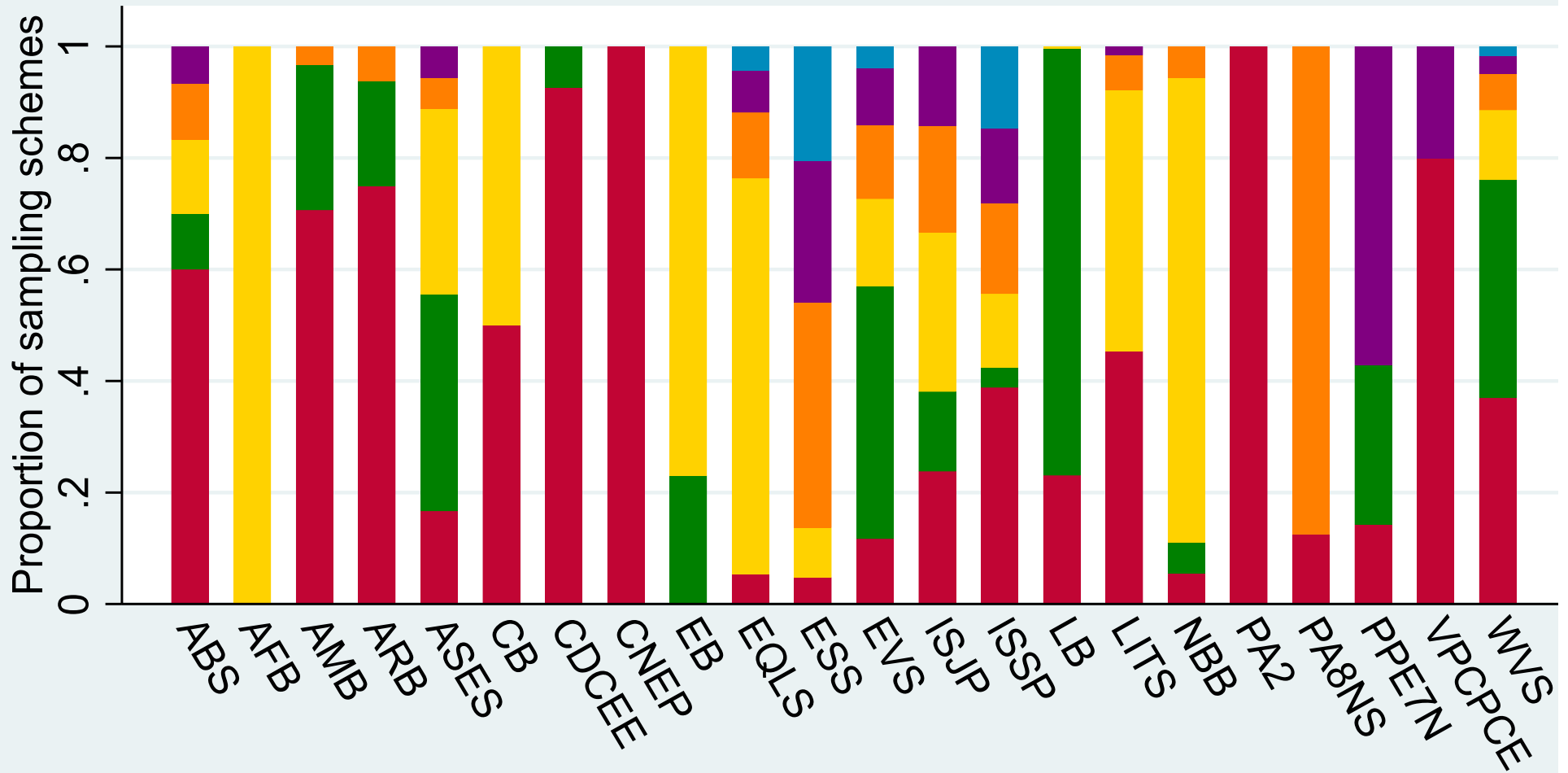
## Changes over time



# Variation across continents



## Variation across projects



Project name	No/insuff info	Quota	Random route	Multi- address	Multi- individual	Single- stage	N surveys
Asian Barometer	60.0	10.0	13.3	10.0	6.7		30
Afrobarometer			100				66
Americas Barometer	70.7	26.1		3.3			92
Arab Barometer	75	18.8		6.3			16
Asia Europe Survey	16.7	38.9	33.3	5.6	5.6		18
Caucasus Barometer	50		50				12
Consolidation of Democracy in CE Europe	92.6	7.4					27
Comparative National Elections Project	100						8
Eurobarometer		23.0	77.0				152
European Quality of Life Survey	5.4		71.0	11.8	7.5	4.3	93
European Social Survey	4.8		8.9	40.4	25.3	20.6	146
European Values Study	11.7	45.3	15.6	13.3	10.2	3.9	128
International Social Justice Project	23.8	14.3	28.6	19.1	14.3		21
International Social Survey Programme	38.8	3.6	13.2	16.3	13.5	14.6	363
Latinobarometro	23.1	76.5	0.4				260
Life in Transition Survey	45.3		46.9	6.3	1.6		64
New Baltic Barometer	5.6	5.7	83.3	5.6			18
Political Action II	100						3
Political Action - An Eight Nation Study	12.5			87.5			8
Political Participation and Equality in 7 Nations	14.3	28.6			57.1		7
Values and Political Change in Postcomm Europe	80				20		5
World Values Survey	37.0	39.13	12.5	6.52	3.26	1.63	184
	<b>27.7</b>	<b>24.5</b>	<b>24.5</b>	<b>10.6</b>	<b>7.2</b>	<b>5.5</b>	<b>1721</b>

# 3 dimensions of Survey Quality

## **Data records: 0(1)4**

- No duplicated cases (0/1)
- Case IDs are non-missing and unique (0/1)
- Item non-response on gender and age below 5% (0/1)
- Weights available and without formal errors (0/1)

## **Descriptive documentation: 0(1)4**

- Documentation contains information about the **response rate**, and provides information that quality **translation method**, **pretesting**, and **fieldwork control** were carried out

## **Consistency between data and documentation: 0/1**

- No discrepancies in values, value labels, variable labels, between the data records and the codebooks



# Sampling and Survey Quality

Sampling	N surveys	Q data 0-4	Q docu 0-4	Q data-docu 0-1
		Mean quality score		
No/insufficient info	477	3.241	1.212	0.518
Quota	422	3.218	0.559	0.301
Random route	421	3.632	1.995	0.458
Multi-stage address	182	3.566	2.863	0.648
Multi-stage individual	124	3.298	2.992	0.629
Single-stage	95	3.137	2.968	0.789
<b>Total</b>	<b>1721</b>	<b>3.364</b>	<b>1.643</b>	<b>0.487</b>

# Sampling and Data Quality

<b>Sampling</b>	<b>Item non-response &lt;5%</b> 0-1	<b>No duplicates</b> 0-1	<b>Unique case IDs</b> 0-1	<b>Correct weights</b> 0-1
Proportion of surveys with positive scores				
<b>No/insufficient info</b>	0.981	0.883	0.964	0.413
<b>Quota</b>	0.981	0.870	0.903	0.464
<b>Random route</b>	0.979	0.931	0.962	0.760
<b>Multi-stage address</b>	0.989	0.962	0.962	0.654
<b>Multi-stage individual</b>	0.992	0.927	0.984	0.395
<b>Single-stage</b>	0.968	0.937	0.989	0.242
<b>Total</b>	<b>0.981</b>	<b>0.906</b>	<b>0.951</b>	<b>0.525</b>

# Sampling and Documentation Quality

---

<b>Sampling</b>	<b>Response rate</b> 0-1	<b>Translation</b> 0-1	<b>Pretest</b> 0-1	<b>Fieldwork control</b> 0-1
	Proportion of surveys providing information			
<b>No/insufficient info</b>	0.400	0.283	0.237	0.291
<b>Quota</b>	0.140	0.111	0.130	0.178
<b>Random route</b>	0.599	0.686	0.361	0.349
<b>Multi-stage address</b>	0.890	0.648	0.588	0.736
<b>Multi-stage individual</b>	0.976	0.758	0.565	0.694
<b>Single-stage</b>	0.979	0.737	0.526	0.726
<b>Total</b>	<b>0.510</b>	<b>0.438</b>	<b>0.318</b>	<b>0.378</b>

---



# Acknowledgements

The SDR project “Survey Data Recycling: New Analytic Framework, Integrated Database, and Tools for Cross-national Social, Behavioral and Economic Research”, is funded by the National Science Foundation, USA (PTE Federal award 1738502). The project runs from September 2017 to August 31, 2021. The SDR project builds on the grant funded by the National Science Centre, Poland (2012/06/M/HS6/00322).

SDR Team at the Polish Academy of Sciences and The Ohio State University: Kazimierz M. Słomczyński (PI), Irina Tomescu-Dubrow (PI), J. Craig Jenkins (PI), Marta Kołczyńska, Ilona Wysmułek, Przemek Powąłko, Nika Palaguta, Weronika Boruc, Denys Lavryk, Marcin W. Zieliński, Bashir Tofangsazi

[dataharmonization.org](https://dataharmonization.org)

<https://dataharmonization.org/newsletter/>