Advances in Comparative Survey Research and Survey Data Harmonization

Marta Kołczyńska

Institute of Political Studies of the Polish Academy of Sciences

RECSM Summer Methods School 2024

Goals of this course

(1) Provide a conceptual overview of the methodological issues involved in comparative analysis of survey data,

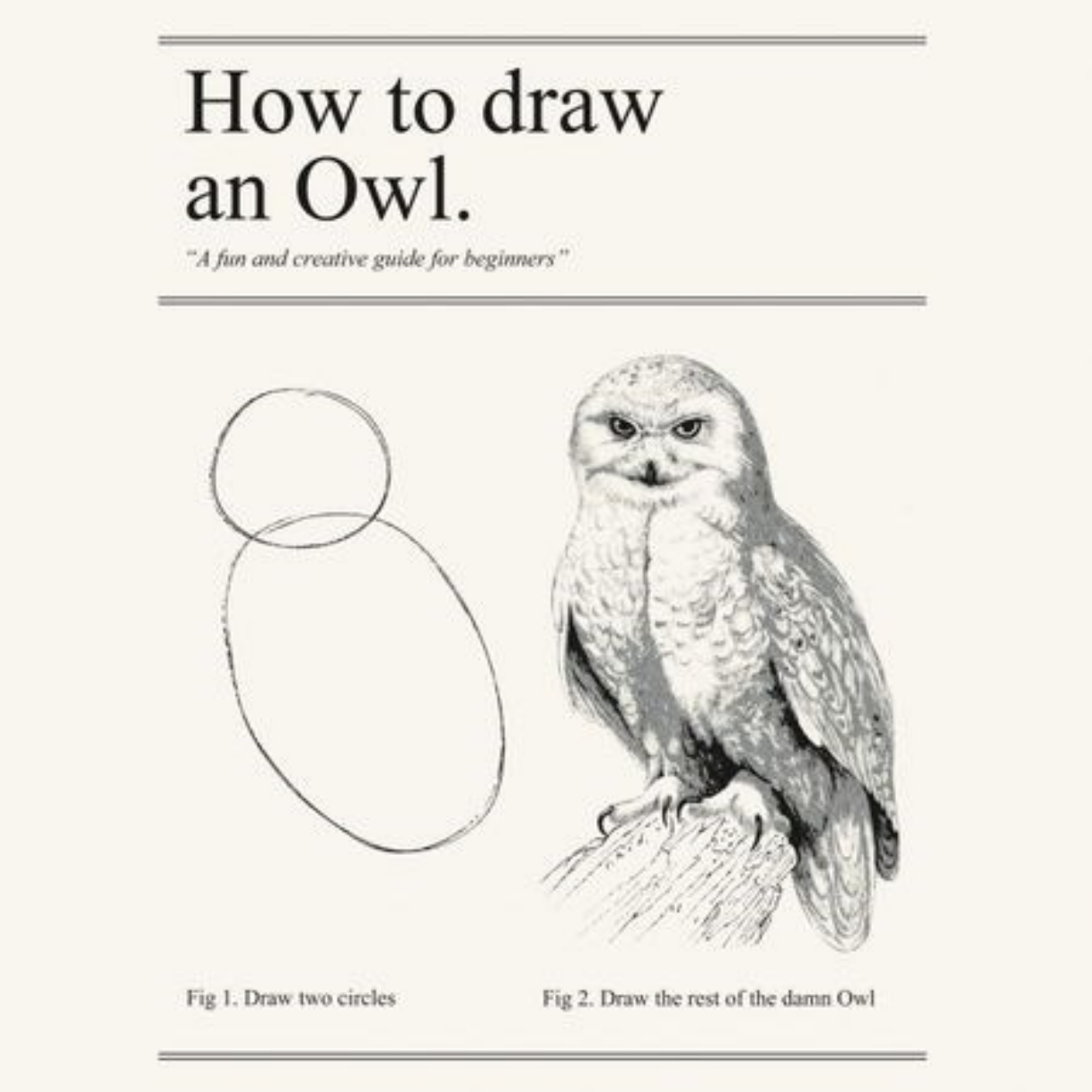
(2) Present ex-post survey data harmonization (integration of existing survey datasets) as a fruitful research strategy,

(3) Present challenges of survey data harmonization and ways of addressing them.

Acknowledgements

NATIONAL SCIENCE CENTRE
POLAND

The content was developed in course of the project *Causes and Consequences of Political Trust: Polarization and Democratic Utility of Trust in Cross-national Perspective* funded in the Sonatina program (2019/32/C/HS6/00421).



https://i0.wp.com/seths.blog/wp-content/uploads/2014/01/6a00d83451b31569e2019aff29b7cd970c-450wi.jpg?ssl=1,
Idea from: Richard McElreath's Bayesian statistics lecture „Statistical Rethinking",
https://www.youtube.com/watch?v=p7g-CgGCS34&ab_channel=RichardMcElreath

Course outline

Day 1

1. Cross-national surveys: overview of available data sources

2. Survey data quality and comparability: Total Survey Error framework, cross-survey differences in measurement and representation

Day 2

3. Framework for survey data harmonization: representativeness and measurement

4. Representation comparability

Day 3

5. Measurement comparability

6. Wrap-up

Breaks: 15.45-16.15 → cafeteria

Perspective: secondary data user

- Access only available materials – at the mercy of data producers (sometimes from a long time ago)

- Restricts the options for analysing data quality

Survey data harmonization

Ex ante harmonization

- Before data collection
- Within survey projects, limited across projects (e.g. borrowing questions)
- Some national statistics across countries
- High effort (planning, organization), high benefit
- Only applicable to future data collections
- Cross-national surveys, e.g. European Social Survey, were ex-ante harmonized (as users, we typically don't think about this)

Ex post harmonization

- After data collection
- Typically by users unrelated to the data collection teams
- Across survey projects
- High effort (data processing, statistics), limited benefit
- The only feasible strategy with historical data

Ex post survey data harmonization

Ex post Survey Data Harmonization is applied to survey datasets that were not *a priori* designed with comparability in mind,

includes procedures that evaluate the quality and comparability of these datasets,

methods of processing the source datasets,

and approaches to analyzing them to achieve research goals.

Ex post survey data harmonization

New interdisciplinary field of study and active area of research (entailing opportunities and pitfalls).

Survey methodology

Subject matter expertise

Computer science / programming skills

Fun fact: studies that do survey data harmonization often don't call it harmonization.

For a historical overview of survey data harmonization efforts see: Dubrow and Tomescu-Dubrow 2016, doi.org/10.1007/s11135-015-0215-z

Ex post survey data harmonization

To study social problems with cross-national survey data, you have to know a lot (1) about survey methodology, and (2) about other countries.

Survey data harmonization adds another layer of complexity but also makes issues inherent in analysis of comparative survey data more salient.

Ex post survey data harmonization

Legal aspects: most survey datasets come with user agreements and data use conditions, which typically prohibit re-publication of their data or its parts.

Some applications

- Combining survey data from Europe and Latin America to examine trajectories in political trust across different types of regimes

- Comparing trends in political trust across European countries since 1990

- Examining macro-level consequences of public opinion
  - Does people's support for democracy strengthen democracy?
  - Does trust in institutions improve the performance of these institutions?

- Examining determinants of public opinion
  - Does the electoral success of populist radical-right parties affect mass attitudes towards immigration?

Political trust in Europe, 1989-2019

There is no single data source that provides enough data for many European countries to reliably estimate trends in political trust.

One can do this with data from 12 cross-national survey projects.

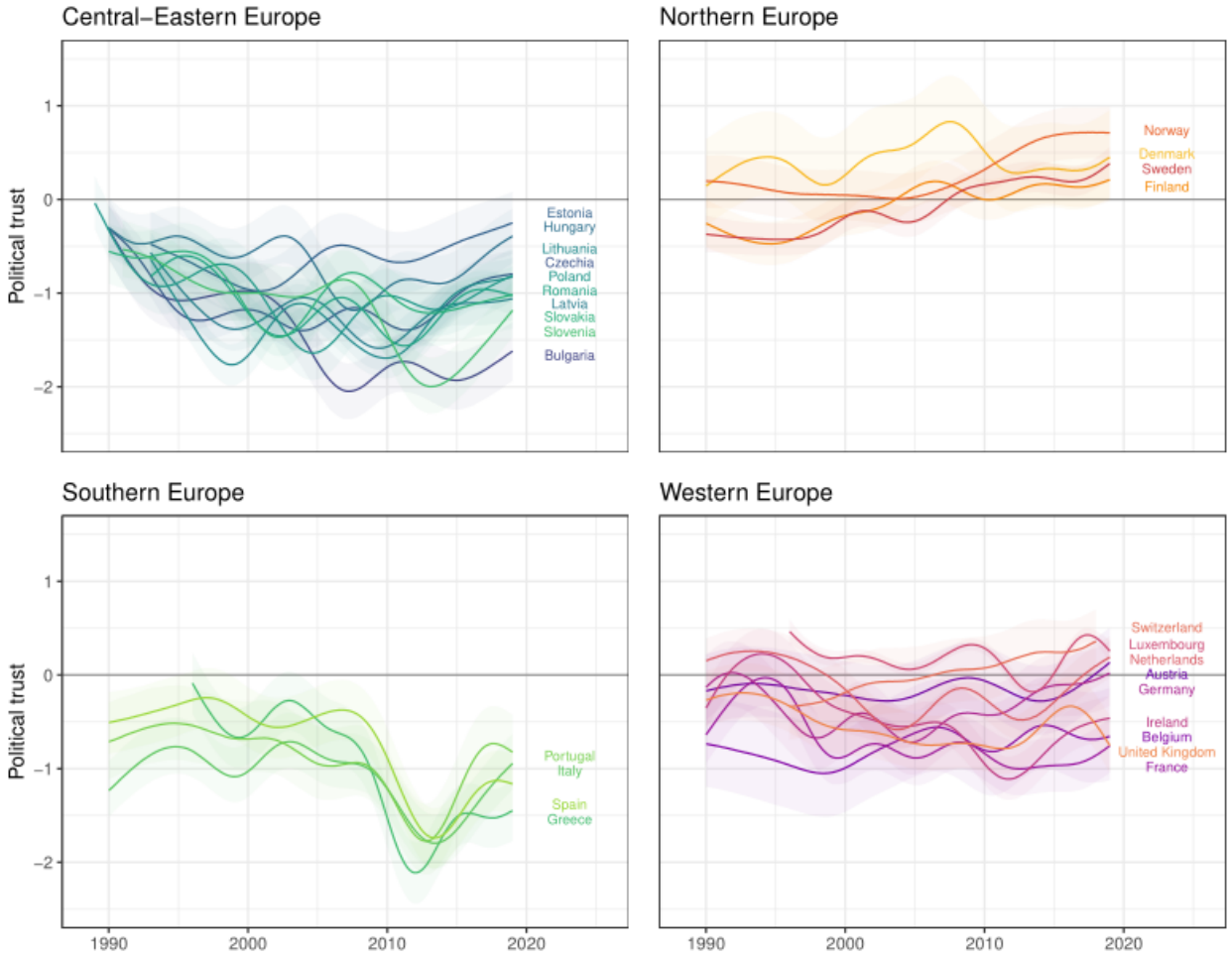


Figure 2: Poststratified estimates of overall levels of political trust by region: posterior medians and 95% credible intervals.

with Paul-Christian Bürkner, Lauren Kennedy, and Aki Vehtari

https://doi.org/10.18148/srm/2024.v18i1.8119

I. Cross-national surveys

Wealth of survey data out there, ready to be analyzed:

Some large and well-known multi-country multi-wave projects

Many smaller scale comparative projects

Countless one-off surveys

Cross-national survey projects

1. Target entire adult population
2. Multi-topic questionnaire

European Social Survey
European Values Study
World Values Survey

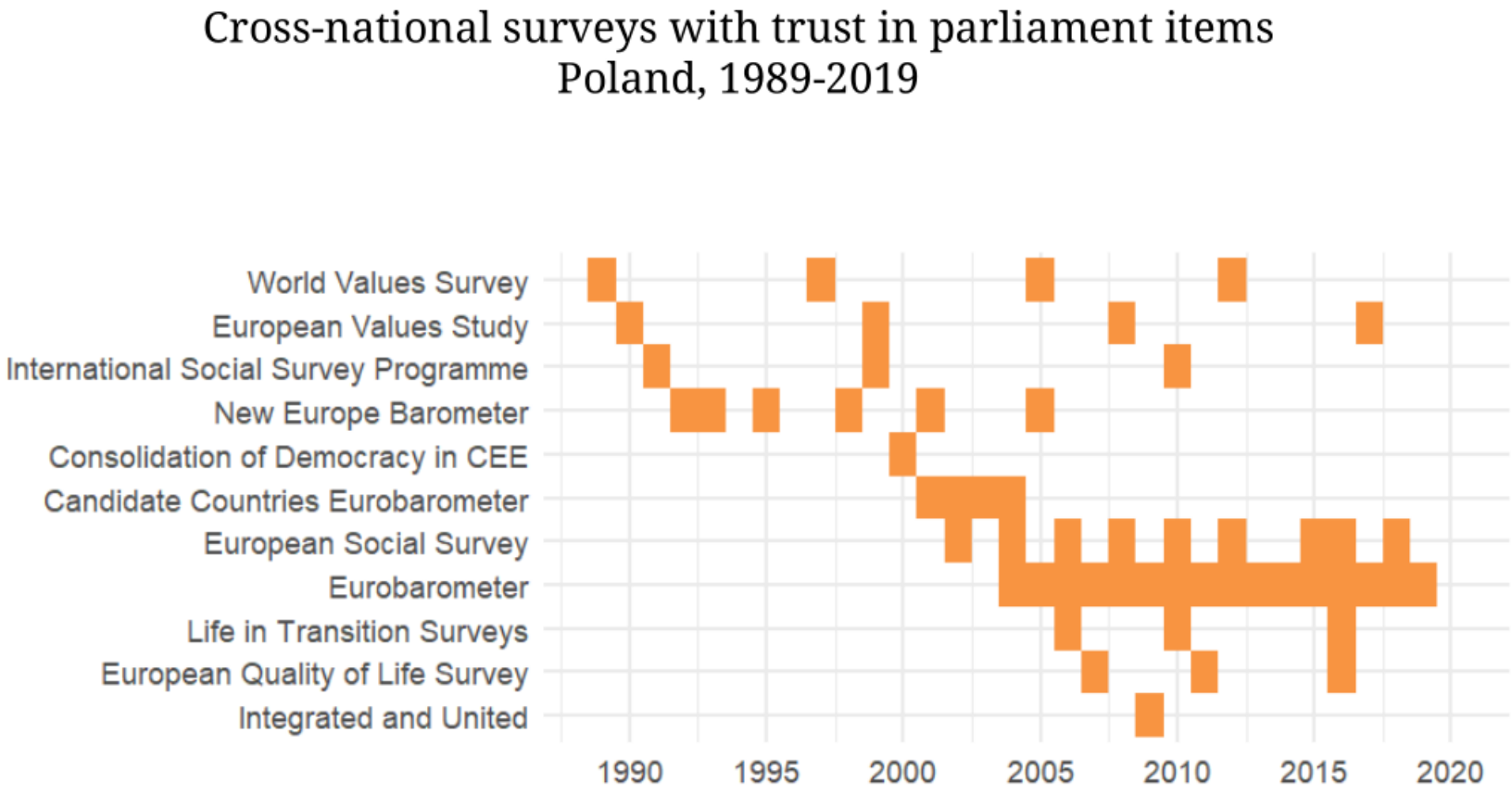
- Projects conducted in post-communist Europe in the 1990s
- Comparative elections studies

Political Action: An 8 Nation Study
Political Action II

Afrobarometer
Arab Barometer
Asian Barometer
Eurasia Barometer
Latinobarometro

Americas Barometer (LAPOP)
Caucasus Barometer
Central Asia Barometer
Eurobarometer
New Europe Barometer

Example: Poland



Alternative: monthly polls from CBOS (Public Opinion Research Centre) covering 1990-2023.

Example: Europe

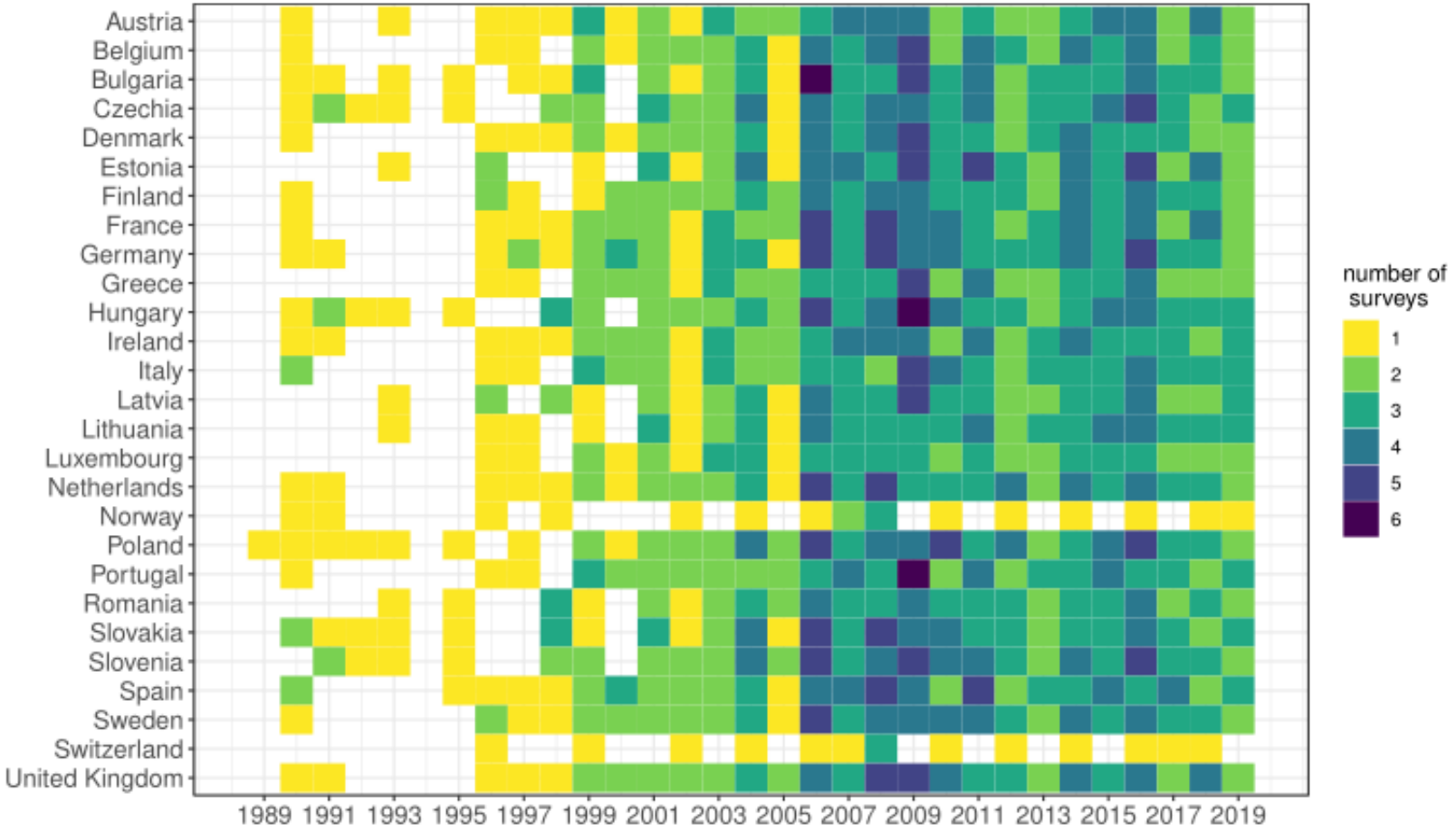


Figure 1: Number of surveys containing any trust item (parliament, parties, justice system) by country and year.

https://osf.io/preprints/socarxiv/3v5g7/

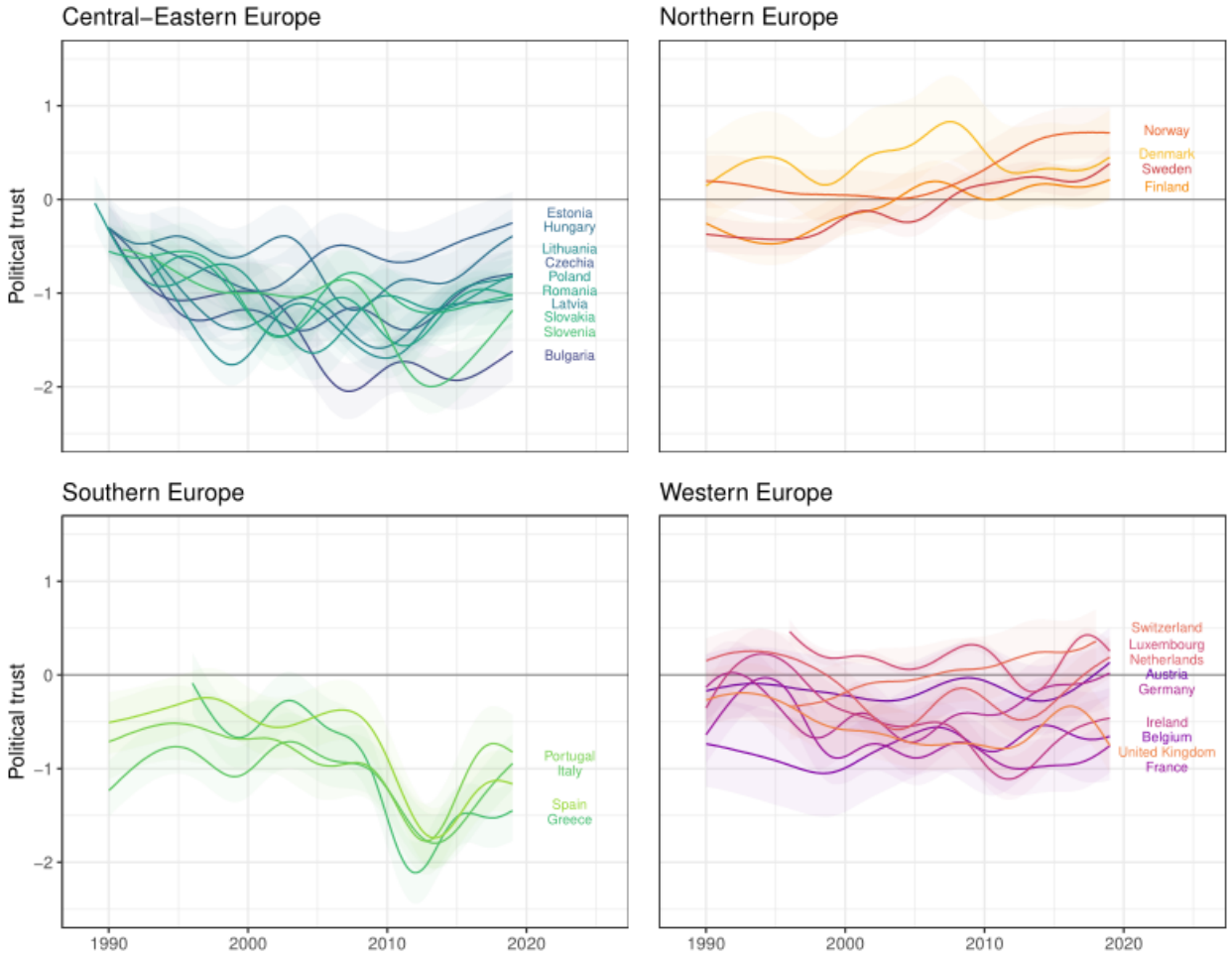


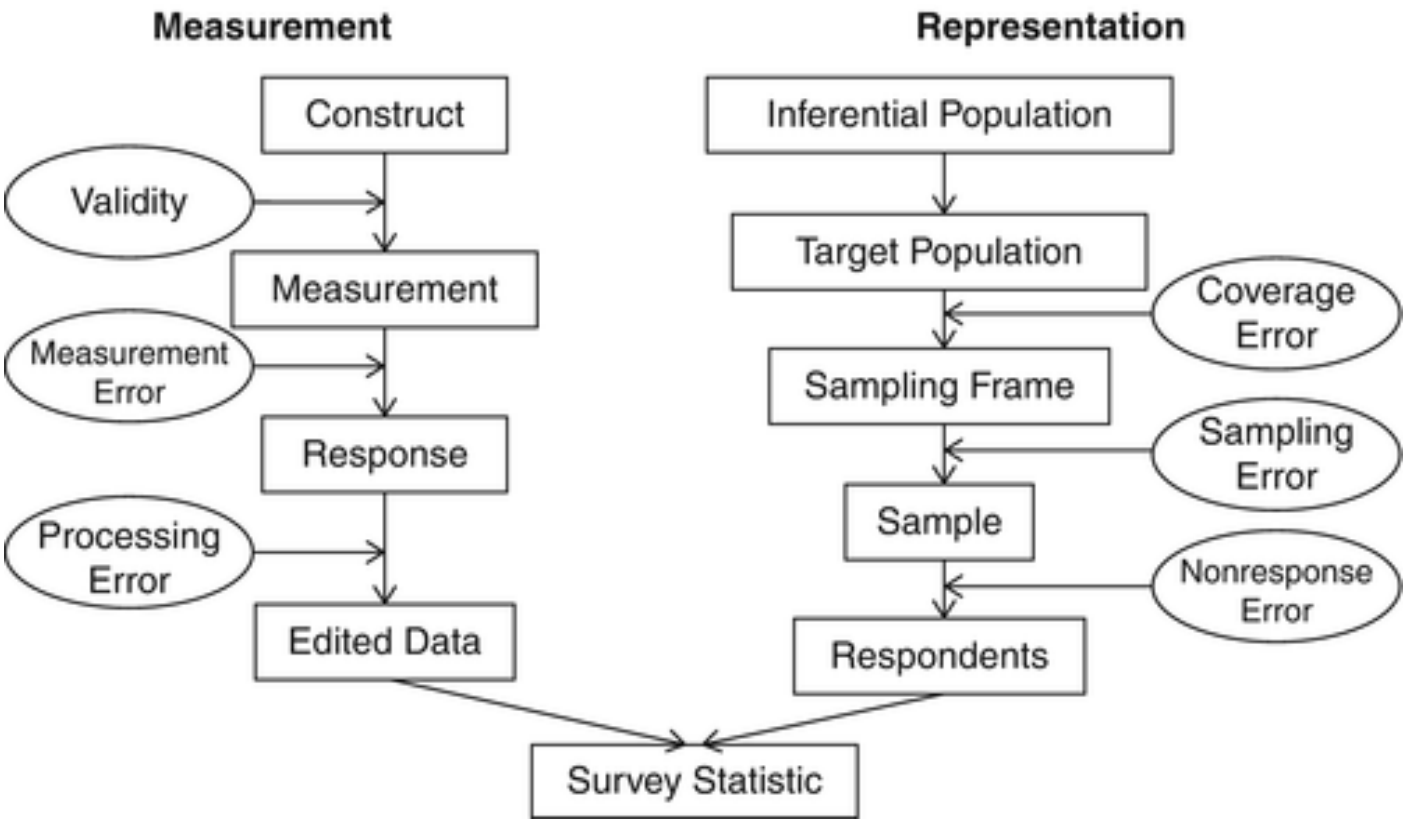
Figure 2: Poststratified estimates of overall levels of political trust by region: posterior medians and 95% credible intervals.

with Paul-Christian Bürkner, Lauren Kennedy, and Aki Vehtari

https://doi.org/10.18148/srm/2024.v18i1.8119

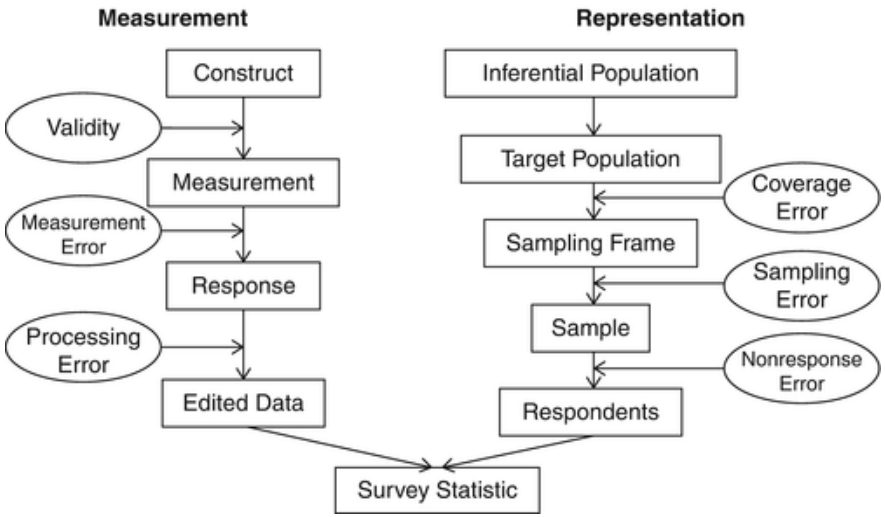
II. Survey data quality and comparability

Total Survey Error framework

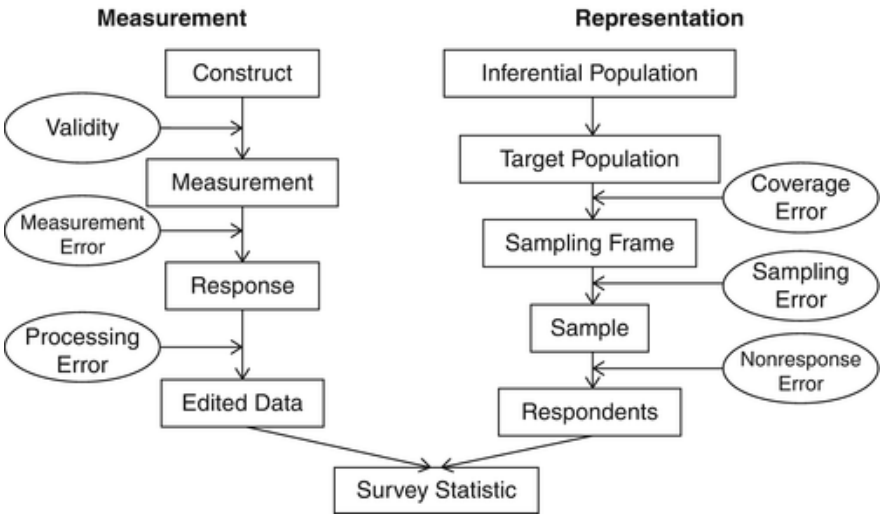


Groves, Fowler, Couper, Lepkowski, Singer, and Tourangeau. 2009. Survey methodology. Wiley.

Survey A



Survey B



Comparability requires similar amounts of errors across surveys.

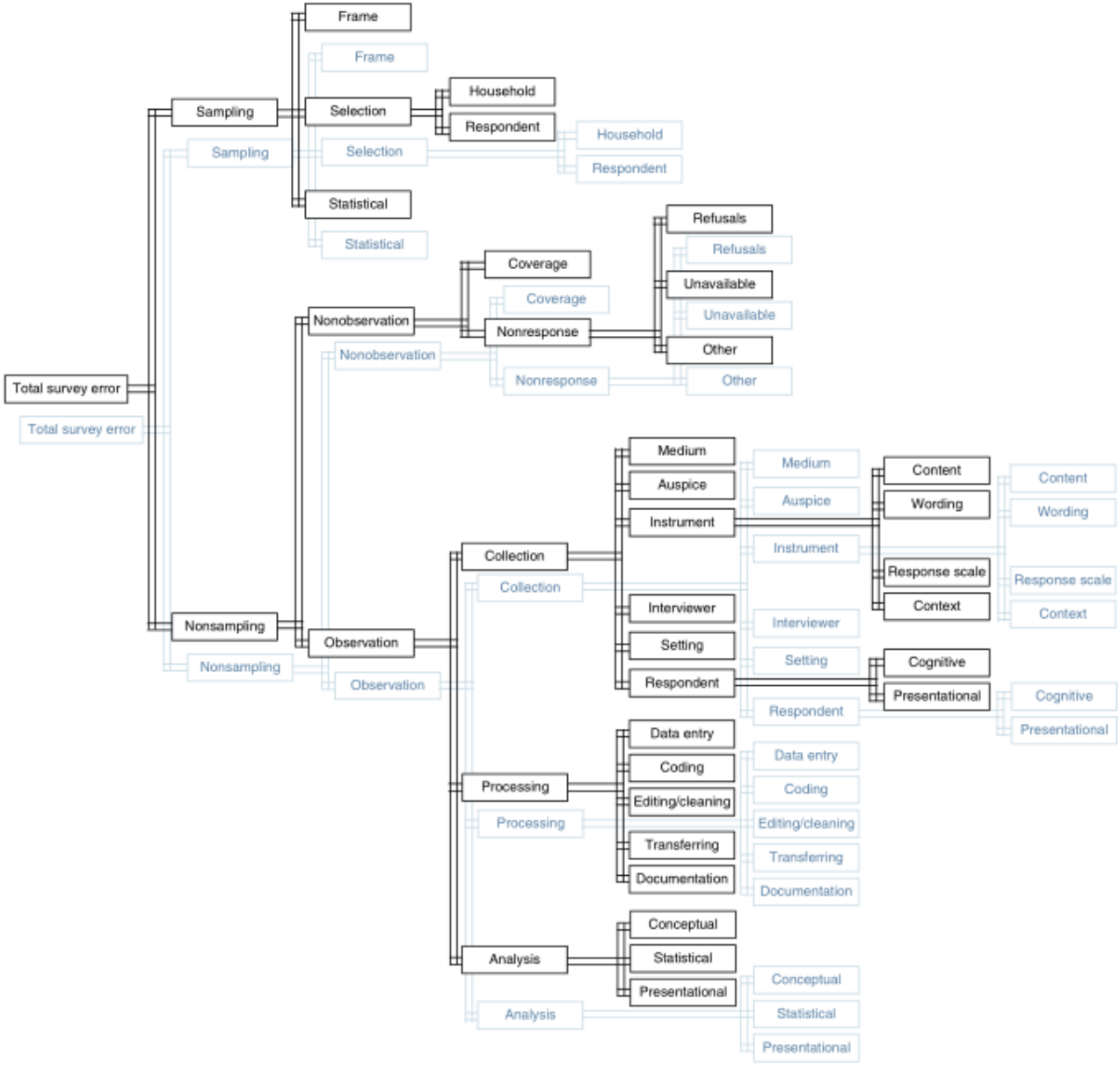


Figure 2.2 Total survey error: Comparison error.

Smith, T.W. (2018). Improving Multinational, Multiregional, and Multicultural (3MC) Comparability Using the Total Survey Error (TSE) Paradigm. In Advances in Comparative Survey Methods (eds T.P. Johnson, B.-E. Pennell, I.A.L. Stoop and B. Dorer). https://doi.org/10.1002/9781118884997.ch2

Comparability

- Should not be assumed
- Needs to be evaluated
- Equivalent vs. identical

Smith, T.W. (2018). Improving Multinational, Multiregional, and Multicultural (3MC) Comparability Using the Total Survey Error (TSE) Paradigm. In Advances in Comparative Survey Methods (eds T.P. Johnson, B.-E. Pennell, I.A.L. Stoop and B. Dorer). https://doi.org/10.1002/9781118884997.ch2

Equivalent vs. Identical

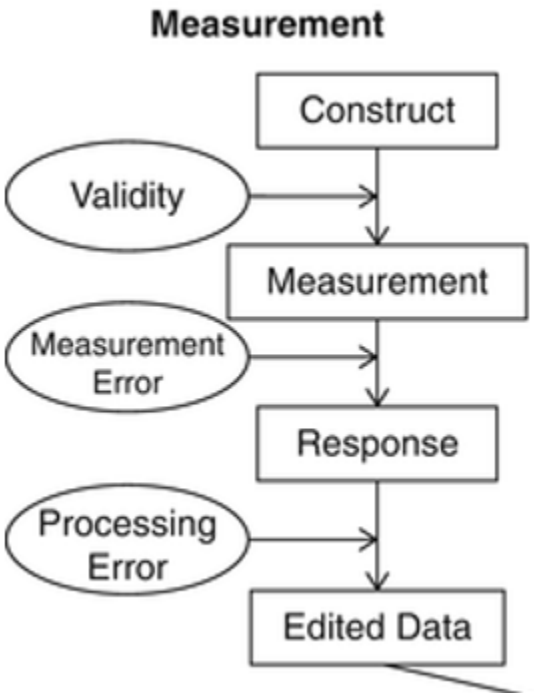
- Different measures/methods may be equivalent
- Identical measures/methods may not be equivalent

Examples:

- Methods of fieldwork control and supervision
- Trust in the healthcare system in Germany and the US
- Satisfaction with the pension system
- Trust in state institutions in democracies and non-democracies

Smith, T.W. (2018). Improving Multinational, Multiregional, and Multicultural (3MC) Comparability Using the Total Survey Error (TSE) Paradigm. In Advances in Comparative Survey Methods (eds T.P. Johnson, B.-E. Pennell, I.A.L. Stoop and B. Dorer). https://doi.org/10.1002/9781118884997.ch2

Measurement




Pieter Bruegel the Elder - The Tower of Babel
Google Art Project
Come, let us go down, and there confound their language, that they may not understand one another's speech.
Genesis 11: 7

Comparability of measurement

It only makes sense to harmonize variables that measure the same thing across surveys.

How to check?

Expert assessment

Literature review

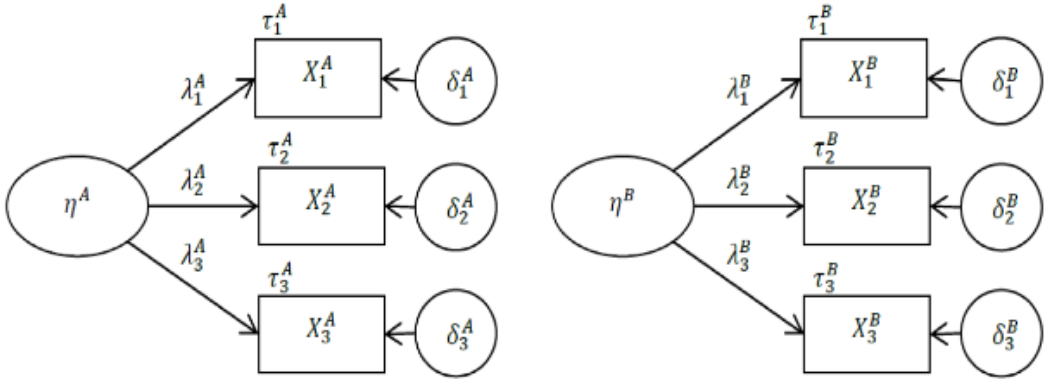
Pretest results

gesis Leibniz Institute for the Social Sciences

by gesis Pretesting Pretest Database

https://pretest.gesis.org/

Multi-Group Confirmatory Factor Analysis (for multi-item scales)



Legend:

$\eta = latent\ variable;\ \lambda = factor\ loading;\ \tau = intercept;\ X = indicator;\ \delta = measurement\ error$

Figure 3.2: Multi-Group Confirmatory Factor Analysis Model

https://bookdown.org/content/5737/invariance.html

Multi-Group Confirmatory Factor Analysis (for multi-item scales)

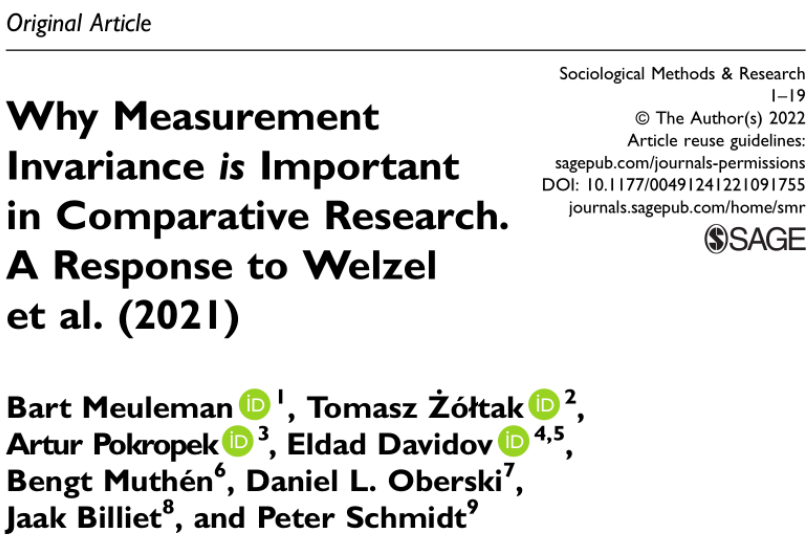
Extensive literature on measurement invariance testing

Strict approaches

Approximate invariance

Alignment method

e.g. Asparouhov and Muthén 2014, doi.org/10.1080/10705511.2014.919210



Original Article

Sociological Methods & Research
1–19

DOI: 10.1177/00491241221091755
journals.sagepub.com/home/smr
SAGE

Why Measurement Invariance *is* Important in Comparative Research. A Response to Welzel et al. (2021)

Bart Meuleman[1], Tomasz Żółtak[2], Artur Pokropek[3], Eldad Davidov[4,5], Bengt Muthén[6], Daniel L. Oberski[7], Jaak Billiet[8], and Peter Schmidt[9]

Validation

Construct validity:

- Convergent validity – expected high correlations with related concepts
- Discriminant validity – expected low correlations with unrelated concepts

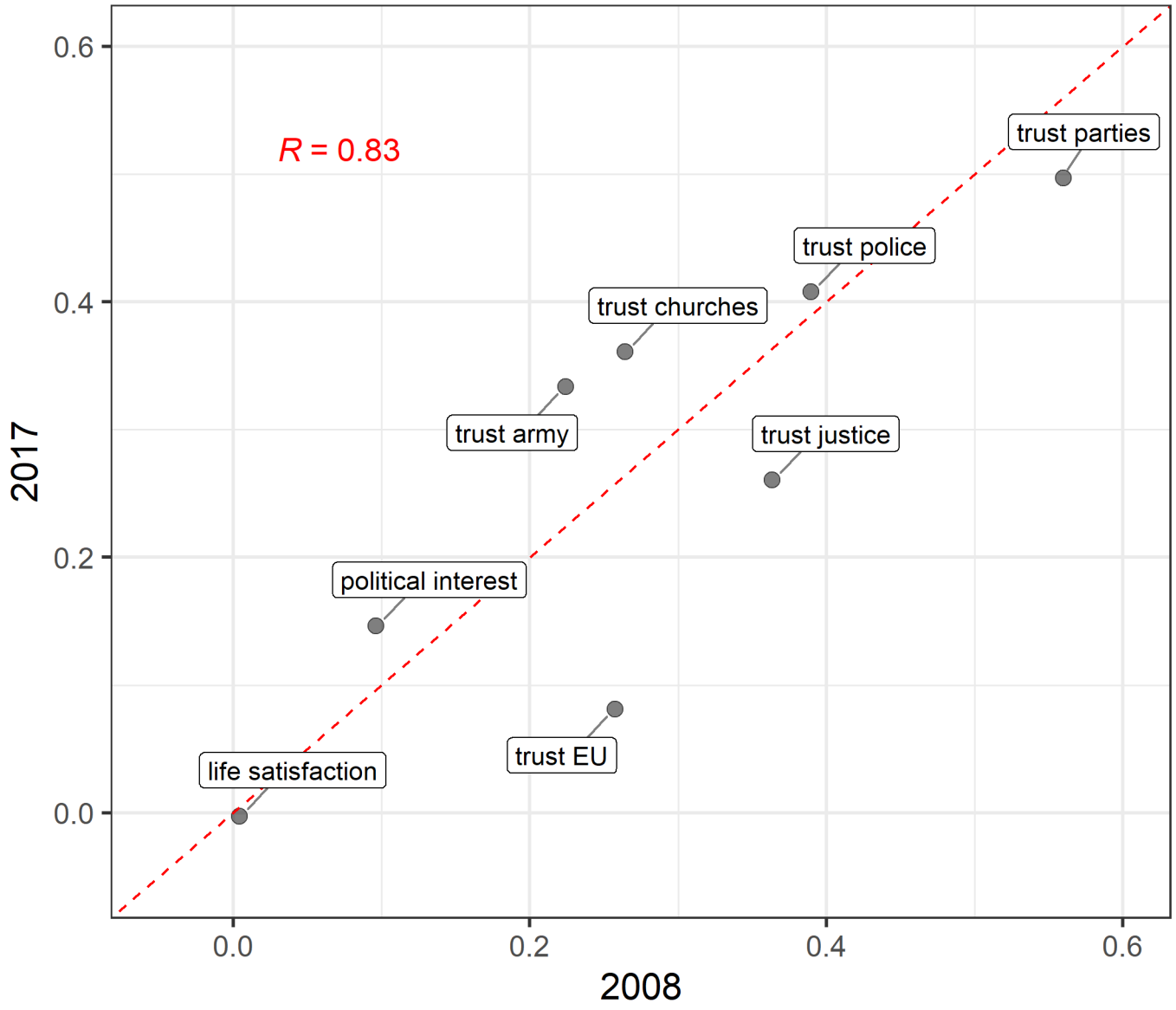
Criterion validity – expected high correlations with outcomes

Example: Poland, EVS 2008, 2017

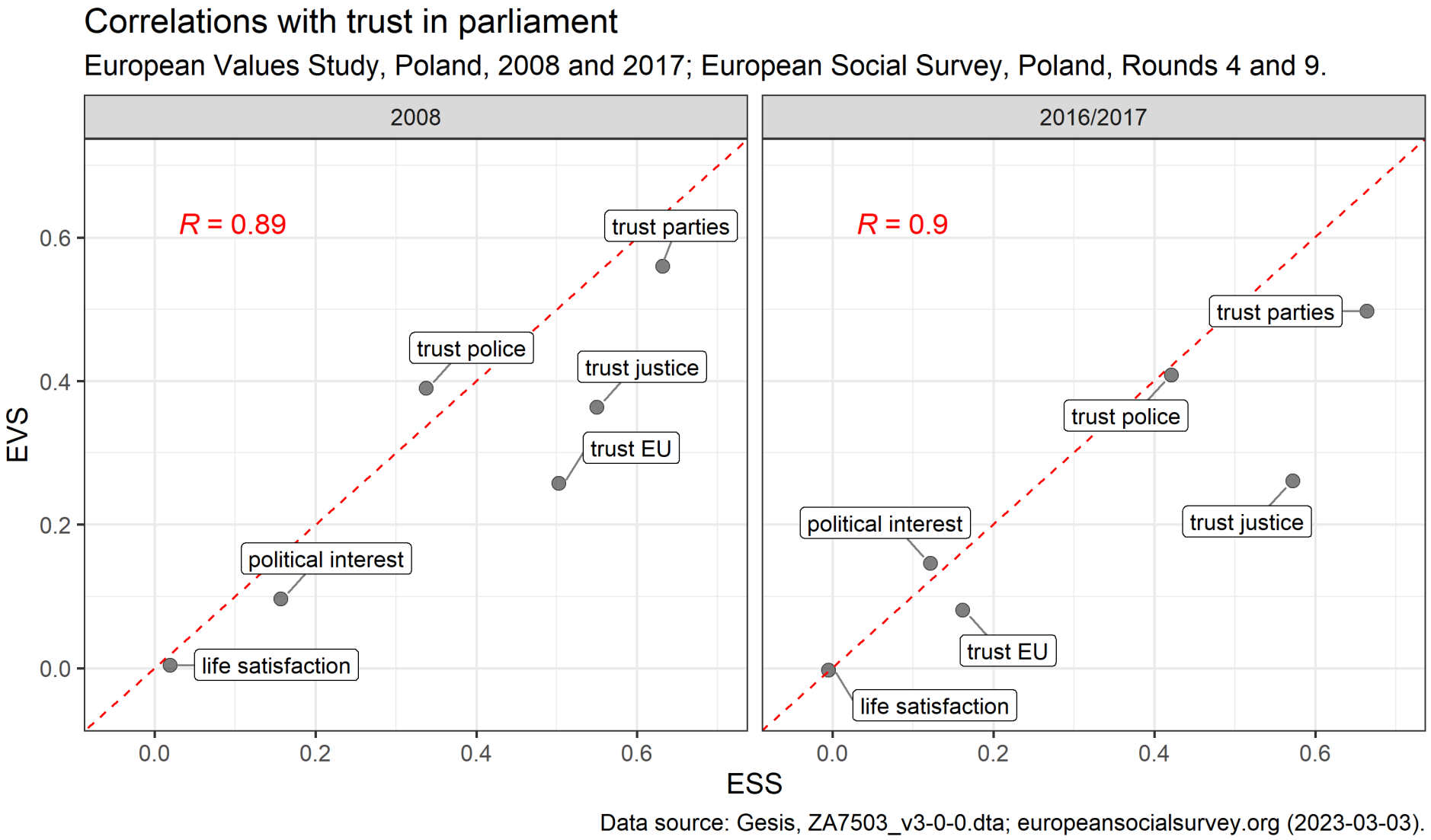
Correlation with trust in parliament:	2008	2017
trust parties	0.560	0.497
trust police	0.390	0.408
trust justice	0.364	0.261
trust churches	0.264	0.361
trust EU	0.258	0.081
trust army	0.224	0.334
political interest	0.096	0.146
life satisfaction	0.004	-0.003

Correlations with trust in parliament

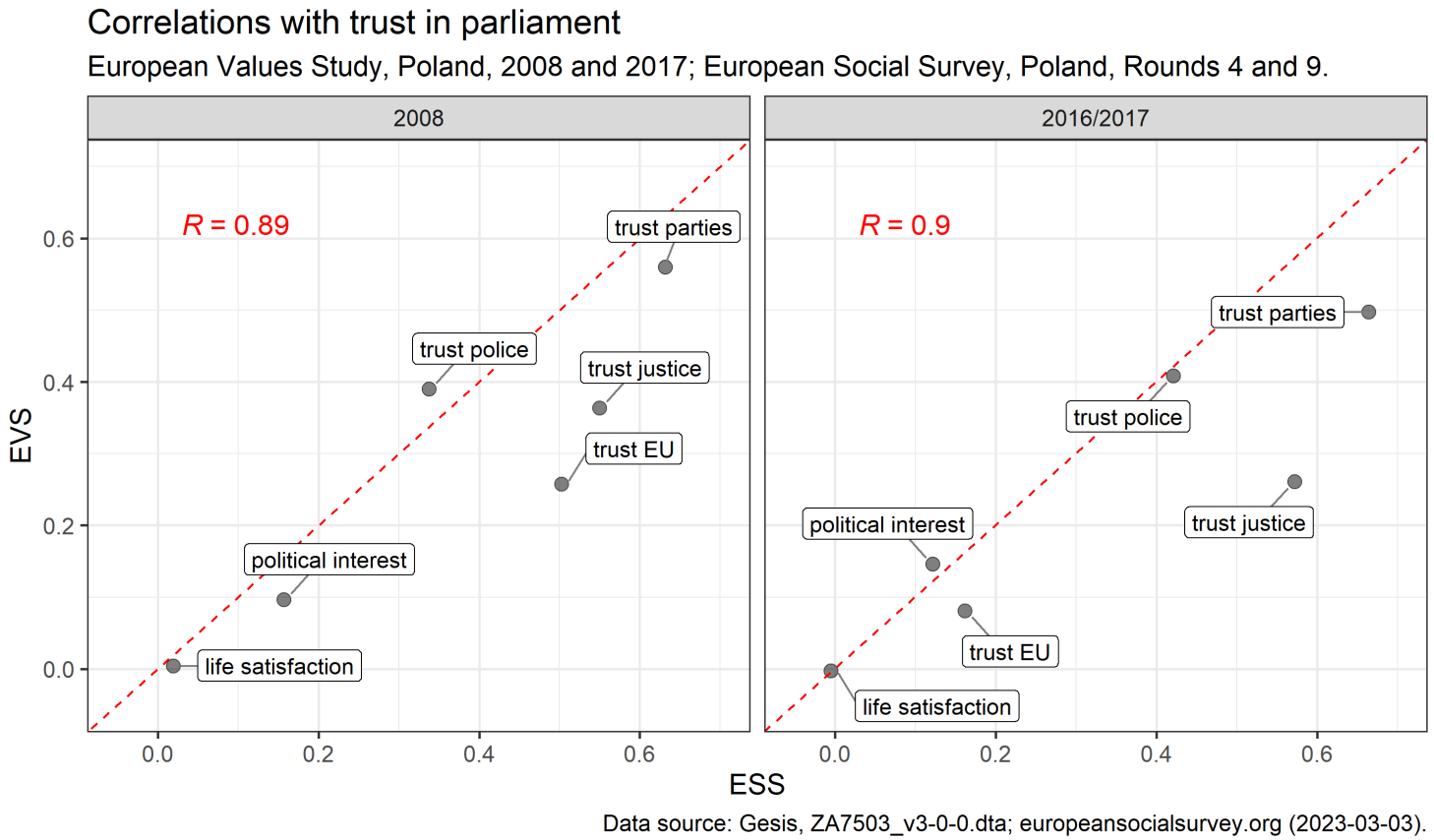
European Values Study, Poland, 2008 and 2017



Data source: https://search.gesis.org/research_data/ZA7503, ZA7503_v3-0-0.dta.



Westen and Rosenthal. 2003. Quantifying construct validity: two simple measures. doi.org/ 10.1037/0022-3514.84.3.608.



Correlations of correlations should be high.

Individual correlations should be close to the 90-degree line rather than form a line of a slope different than 1.

Westen and Rosenthal. 2003. Quantifying construct validity: two simple measures. doi.org/ 10.1037/0022-3514.84.3.608.

Validation helps detect problems

World Values Survey 7

POLITICAL CULTURE & POLITICAL REGIMES

I'm going to describe various types of political systems and ask what you think about each as a way of governing this country. For each one, would you say it is a very good, fairly good, fairly bad or very bad way of governing this country? (*Read out and code one answer for each*):

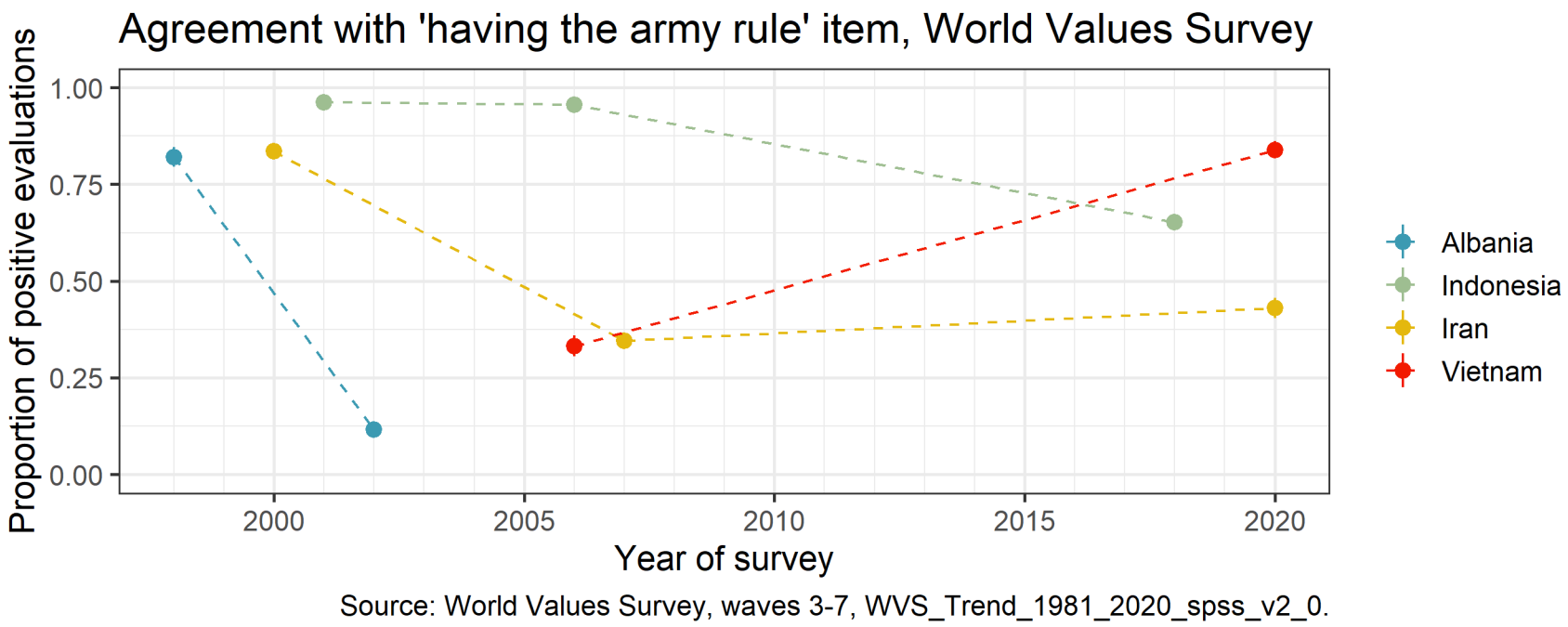
		Very good	Fairly good	Fairly bad	Very bad
Q235	Having a strong leader who does not have to bother with parliament and elections	1	2	3	4
Q236	Having experts, not government, make decisions according to what they think is best for the country	1	2	3	4
Q237	Having the army rule	1	2	3	4
Q238	Having a democratic political system	1	2	3	4
Q239	Having a system governed by religious law in which there are no political parties or elections	1	2	3	4

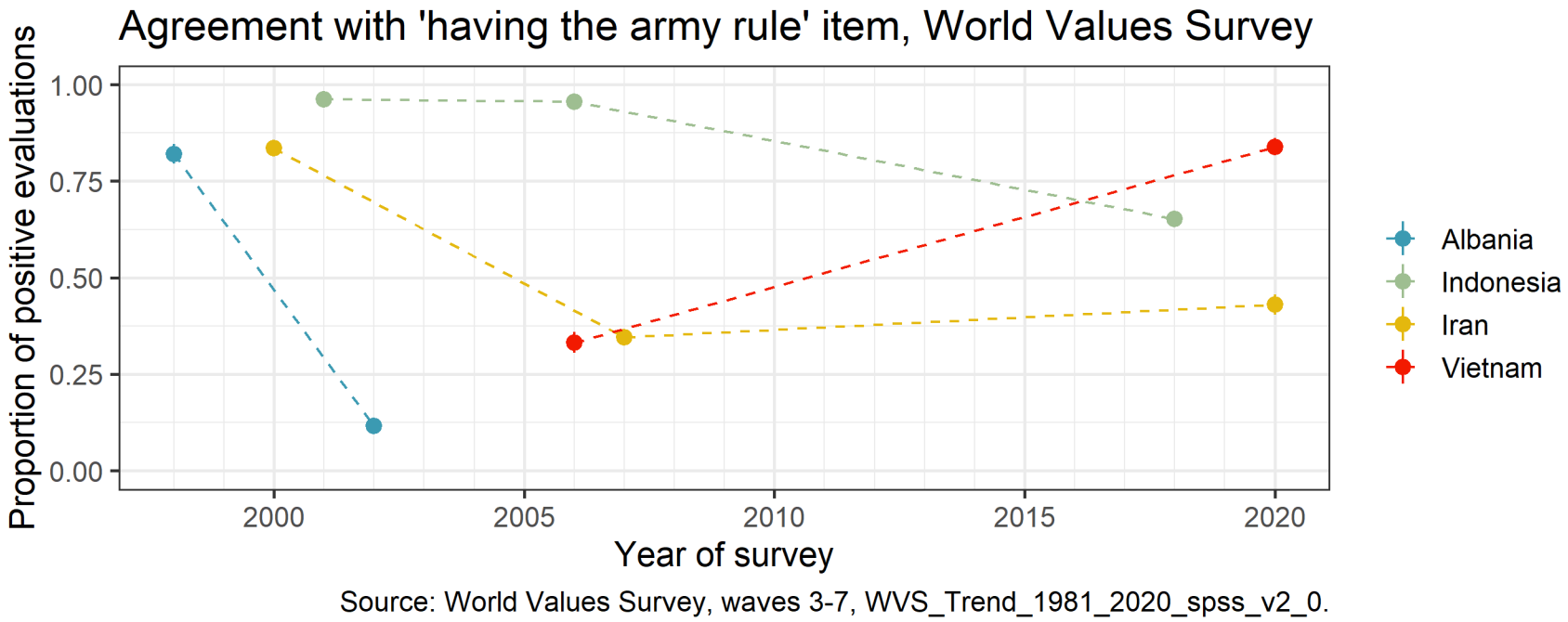
The general coding for missing codes is as follows (do not read them and code only if the respondent mentions them:

-1 Don't know -3 Not applicable (filter)

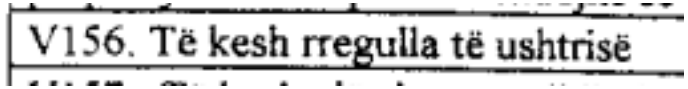
-2 No answer/refused -5 Missing; Not applicable for other reasons

WVS core questionnaire from: https://www.worldvaluessurvey.org/

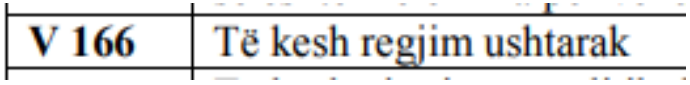

Agreement with 'having the army rule' item, World Values Survey
Proportion of positive evaluations
1.00
0.75
0.50
0.25
0.00
2000
2005
2010
2015
2020
Year of survey
Albania
Indonesia
Iran
Vietnam
Source: World Values Survey, waves 3-7, WVS_Trend_1981_2020_spss_v2_0.



Albanian questionnaires:

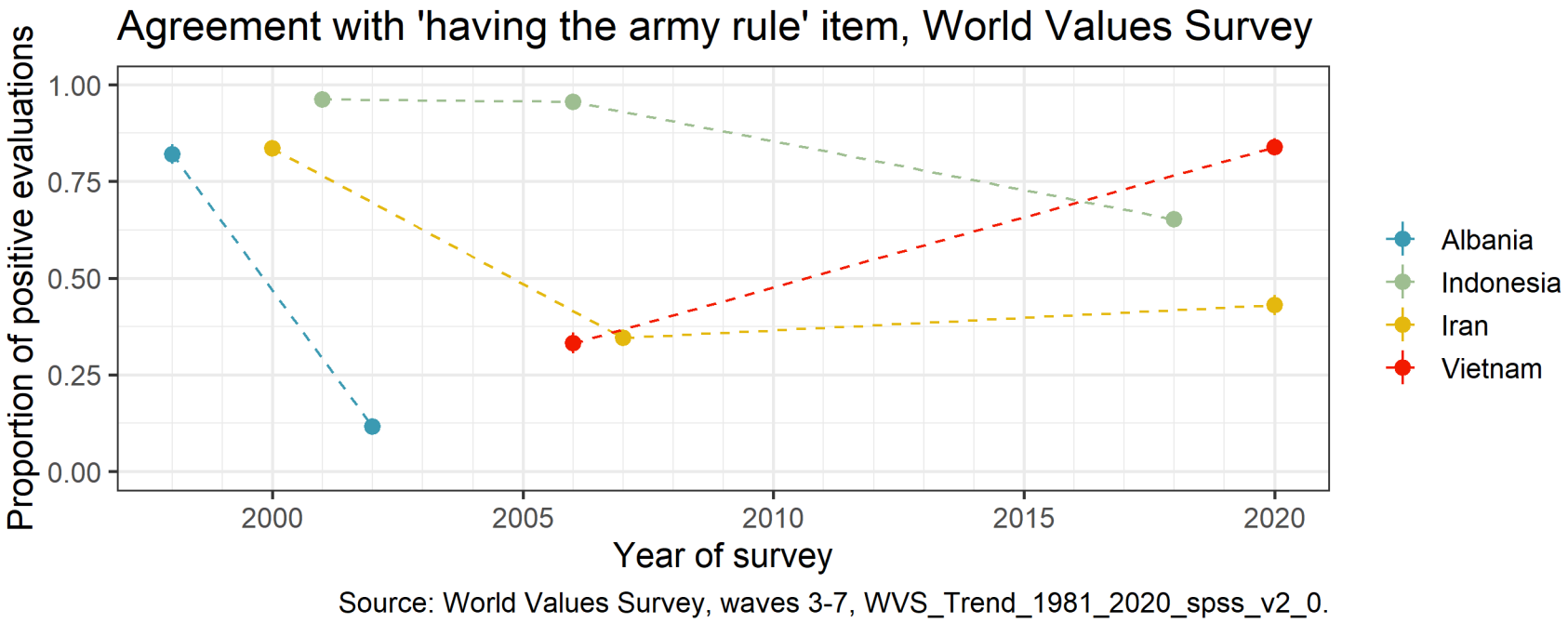


1998: to have military rules



2002: to have a military regime

Story described in: https://www.washingtonpost.com/news/monkey-cage/wp/2014/09/02/world-values-lost-in-translation/; WVS country questionnaires from: https://www.worldvaluessurvey.org/



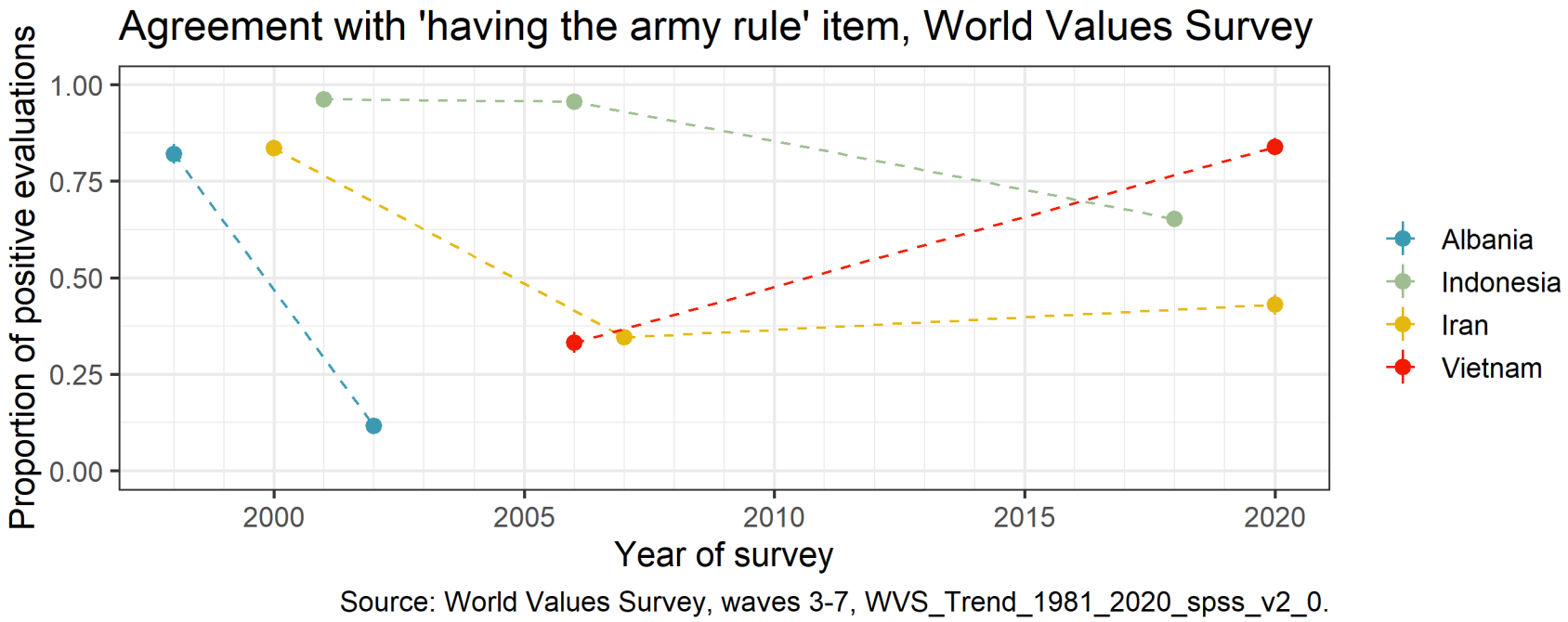
Indonesian questionnaires (google translated):

2001: have clear regulations on the armed forces

2006: have regulations on the armed forces

2018: the army holds the power

WVS country questionnaires from: https://www.worldvaluessurvey.org/



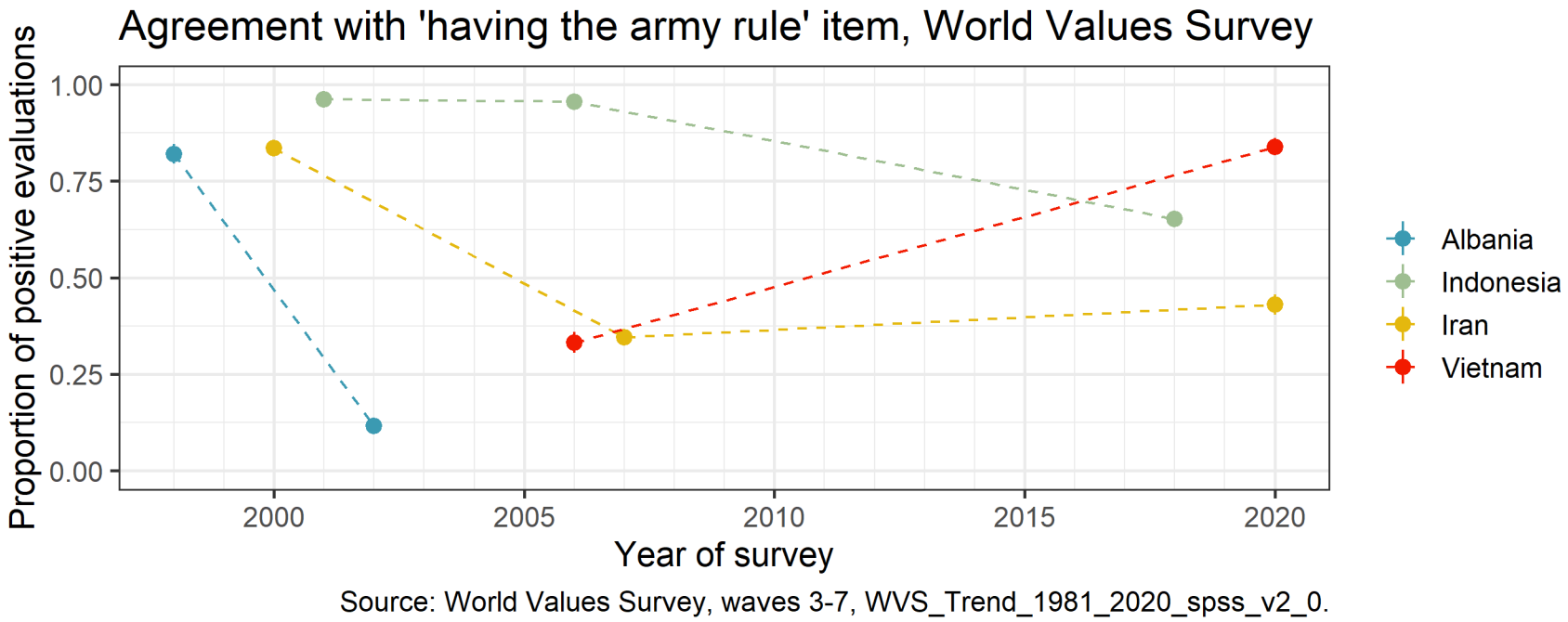
Iranian questionnaires (google translated):

2000: strong government

2007: government military

2020: administration of the country by the army and soldiers

WVS country questionnaires from: https://www.worldvaluessurvey.org/



Vietnamese questionnaires (google translated):

2006: the role of the military

2020: there is military rule

WVS country questionnaires from: https://www.worldvaluessurvey.org/

WVS 5: participation

Have you or have you not done any of these activities in the last five years? (*Read out and code one answer for each action*):

	Have done	Have not done
V100. Signing a petition	1	2
V101. Joining in boycotts	1	2
V102. Attending peaceful demonstrations	1	2
V103. Other (write in):_____	1	2

V98, V102: „Lawful“ instead of „peaceful“ used in Split B, OECD-countries

According to country questionnaires:

in Hong Kong the question asked about the last 12 months,

in Zambia about the last year,

in Jordan there seems to be no indication of the time frame.

doi.org/10.12758/mda.2019.07, fn. 2; WVS questionnaires from: https://www.worldvaluessurvey.org/

ESS: Religiosity scale

C13 *Regardless of whether you belong to a particular religion, how religious would you say you are?*
(0 = not at all; –10 = very religious)

C14 *Apart from special occasions such as weddings and funerals, about how often do you attend religious services nowadays?*
(Reversed: 0 = never; –7 = everyday)

C15 *Apart from when you are at religious services, how often, if at all, do you pray?*
(Reversed: 0 = never; –7 = everyday)

Using this scale of religious involvement, in most countries women are more religious than men, but in Turkey women are much less religious than men.

Meuleman, Bart and Jaak Billiet (2011), 'Religious involvement and its relations to values and social attitudes', in Eldad Davidov, Peter Schmidt and Jaak Billiet (eds), Cross-cultural Analysis: Methods and Applications, pp. 173–206.

Translation and other measurement issues: how to avoid

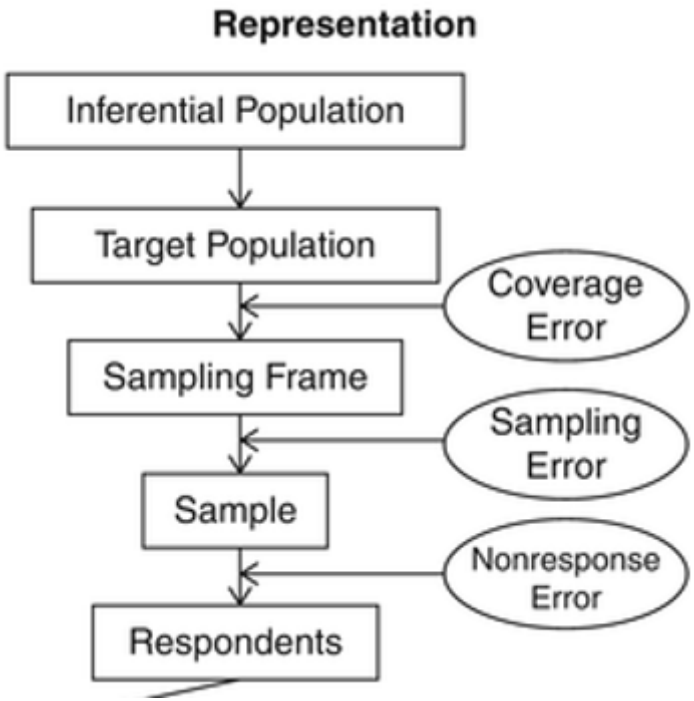
Checking country questionnaires (if available).

The Multilingual Corpus of Survey Questionnaires (https://www.upf.edu/web/mcsq) hosted here at UPF.

Measurement invariance tests for scales.

Triple-check all anomalies.

Representation



Comparability of representation

Target populations: „general adult population samples” often differ in (at least):

- Age
- Nationality / citizenship

Sampling frame: some surveys do not have frames (quota samples, random walk samples without screening)

Nonresponse

Generally, over time, survey coverage has become better, some sample designs have become better (but also non-probability samples become more widespread), but nonresponse has become worse and surveys have become more expensive to conduct.

Target populations

Minimum age:

Eurobarometer, European Social Survey = 15 years

European Quality of Life Survey, most of International Social Survey Programme, European Values Study = 18 years

Some ISSP = 21 years

Some surveys have upper age limits, some as low as 65.

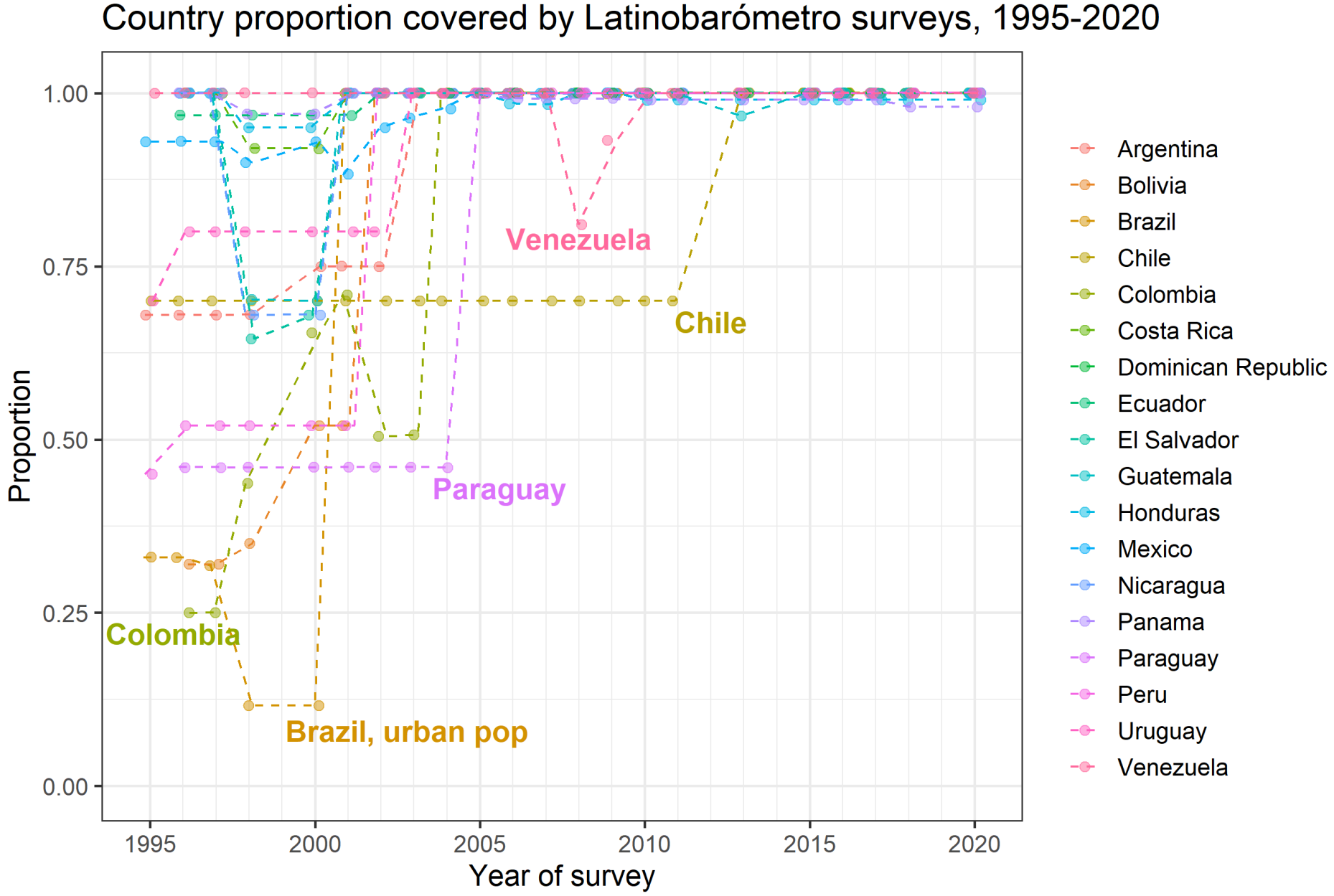
Exclusions based on language, nationality, residence, territorial exclusions, etc.

Latinobarómetro: coverage

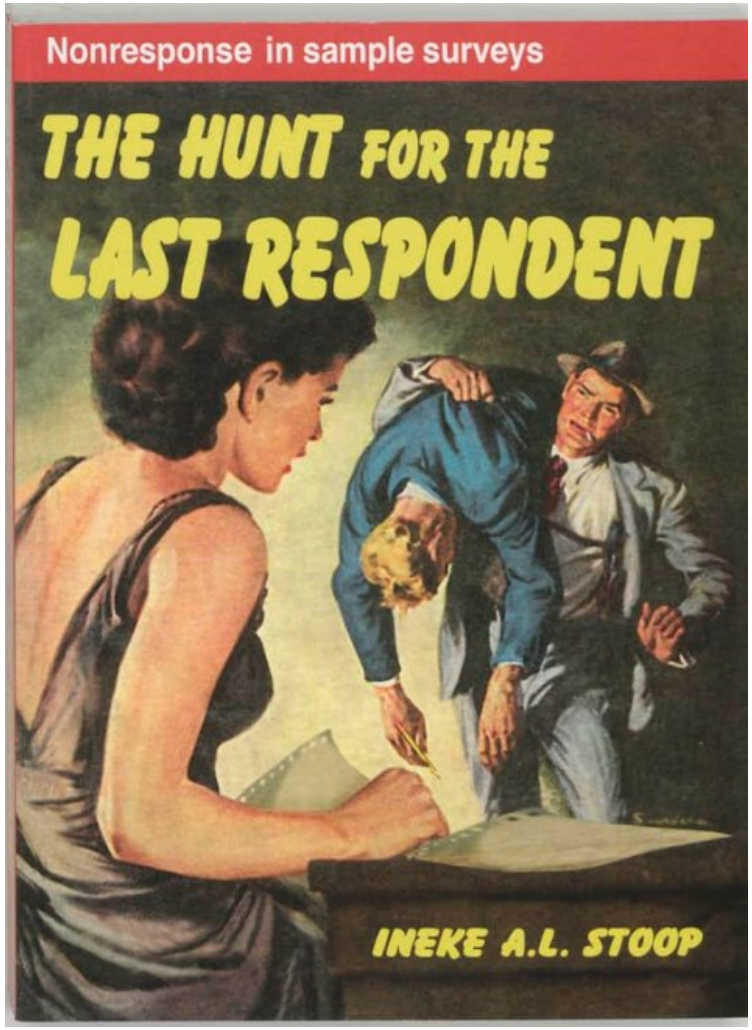
- Annual surveys since 1995 (8 countries) until 2018 (18 countries) + 2020

- The documentation provides the % of country covered by each survey, but does not explain what territories/groups are excluded

https://www.latinobarometro.org/


Country proportion covered by Latinobarómetro surveys, 1995-2020
Proportion
1.00
0.75
0.50
0.25
0.00
1995
2000
2005
2010
2015
2020
Year of survey
Venezuela
Chile
Paraguay
Colombia
Brazil, urban pop
Argentina
Bolivia
Brazil
Chile
Colombia
Costa Rica
Dominican Republic
Ecuador
El Salvador
Guatemala
Honduras
Mexico
Nicaragua
Panama
Paraguay
Peru
Uruguay
Venezuela
Source: latinobarometro.org, technical data sheets 1995-2020.

Response rates decline worldwide



Published in 2005

JOURNAL ARTICLE

Where Have the Respondents Gone? Perhaps We Ate Them All FREE

Thomas J Leeper

Public Opinion Quarterly, Volume 83, Issue S1, 2019, Pages 280–288,
https://doi.org/10.1093/poq/nfz010
Published: 19 June 2019

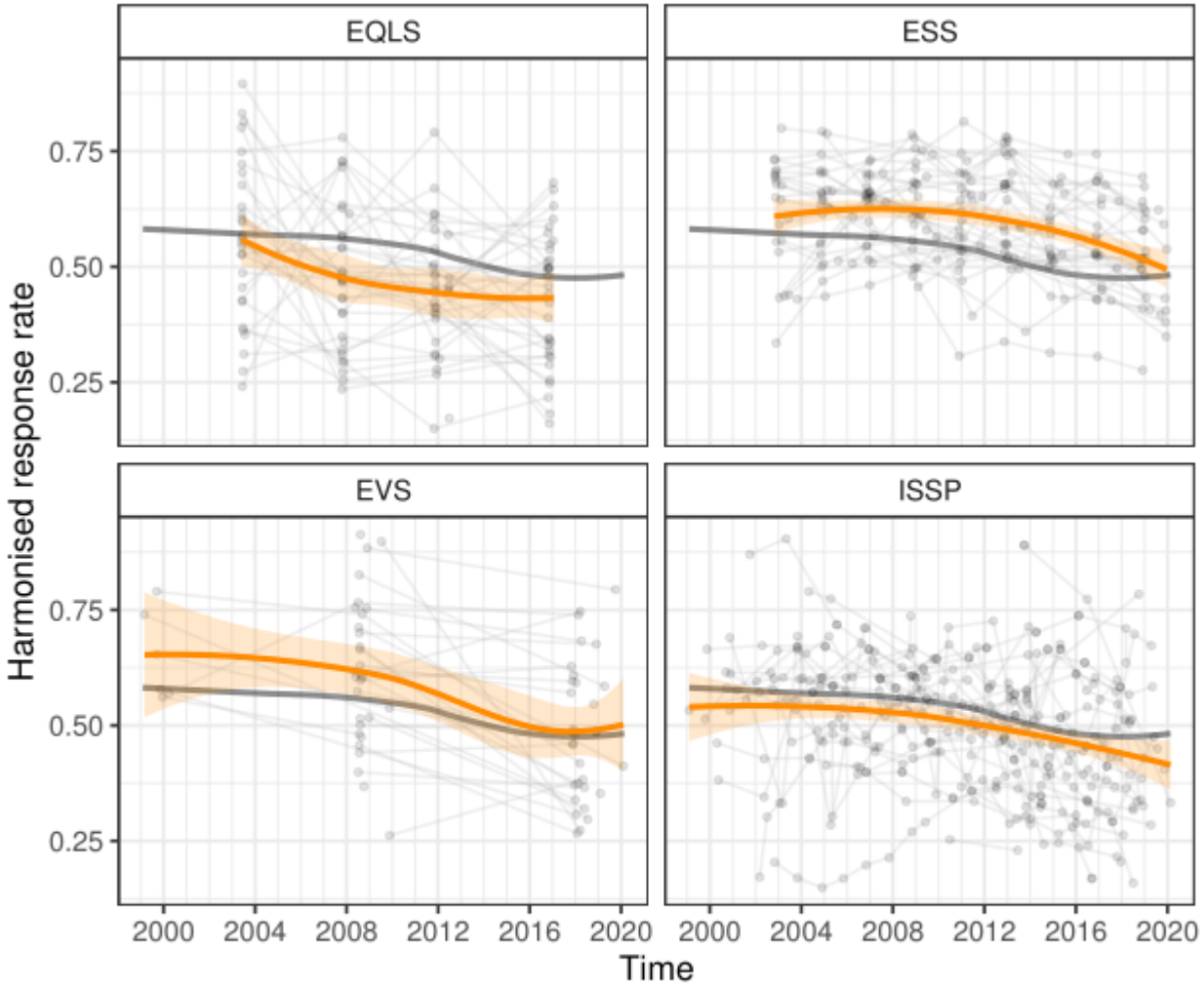


Figure 8. Harmonised response rates over time by project. The grey curve indicates the LOESS trend in all surveys in the estimation sample. The orange curves indicate the LOESS trends within each project, with the 95% confidence interval indicated by the ribbon.

Unpublished draft, Jabkowski, Kohler, and Kołczyńska.

Response rates

- There are different definitions of response rates
- AAPOR's Standard Definitions: https://aapor.org/standards-and-ethics/standard-definitions/

Numerator: Complete interviews (or + partial interviews)

Denominator: Complete + Partial + Refusals + Break-offs + Other
(or + Unknown eligibility + Unknown if HH occupied)

$$RR1 = \frac{Complete\ interviews}{Complete + Partial + Refusals + Breakoffs + Other + Unknown\ elig.}$$

Bias

Nonresponse bias (Bethlehem 1988, 254):

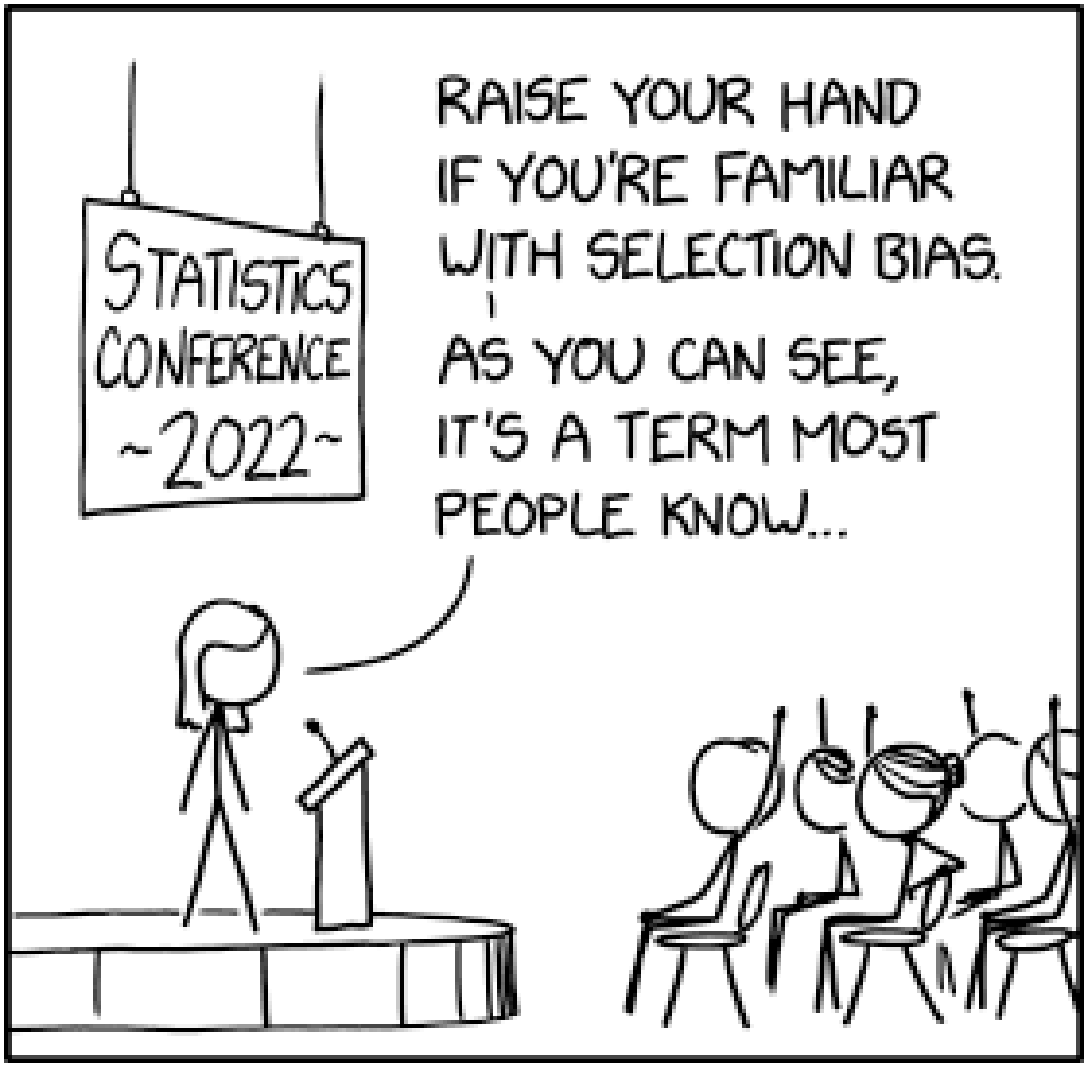
$$bias(\bar{Y}) = \frac{cov(Y, \pi)}{\bar{\pi}}$$

Y – target variable

π – response probabilities

Bias is larger the higher the covariance and the lower the response rate.

See also: Bradley, V. C., Kuriwaki, S., Isakov, M., Sejdinovic, D., Meng, X.-L., & Flaxman, S. (2021). Unrepresentative big surveys significantly overestimated US vaccine up-take. Nature, 600(7890), 695–700. https://doi.org/10.1038/s41586-021-04198-4



https://xkcd.com/2618/

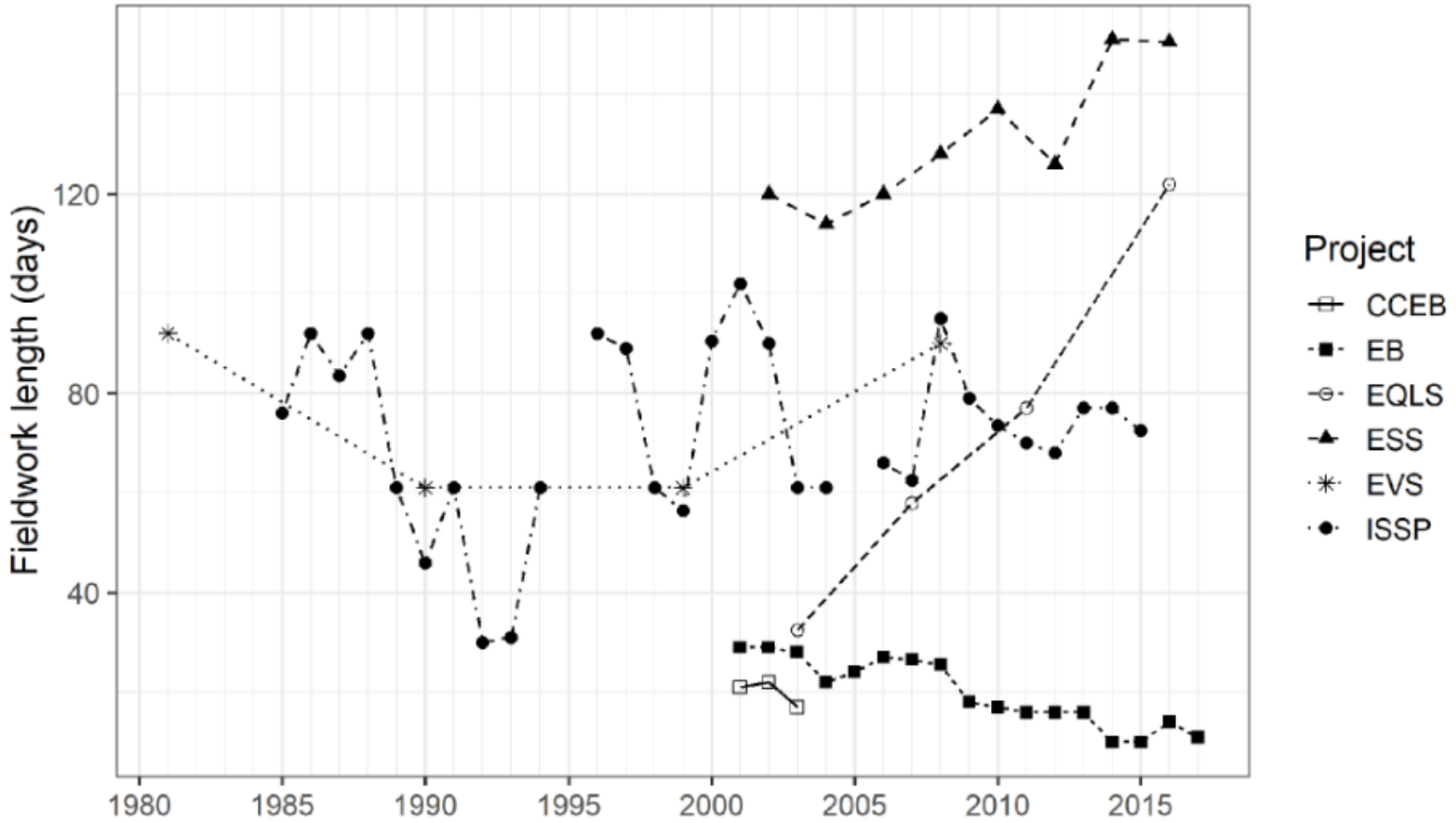
Exercise

ex1_sampling.R from Aula Global

Tidyverse is „an opinionated collection of R packages designed for data science” (www.tidyverse.org).

Figure 2

Median Length of Fieldwork in Days in Different Editions of Survey Projects



https://doi.org/10.5964/meth.2795

Fieldwork length - consequences

Interviewers first reach respondents who are more accessible:

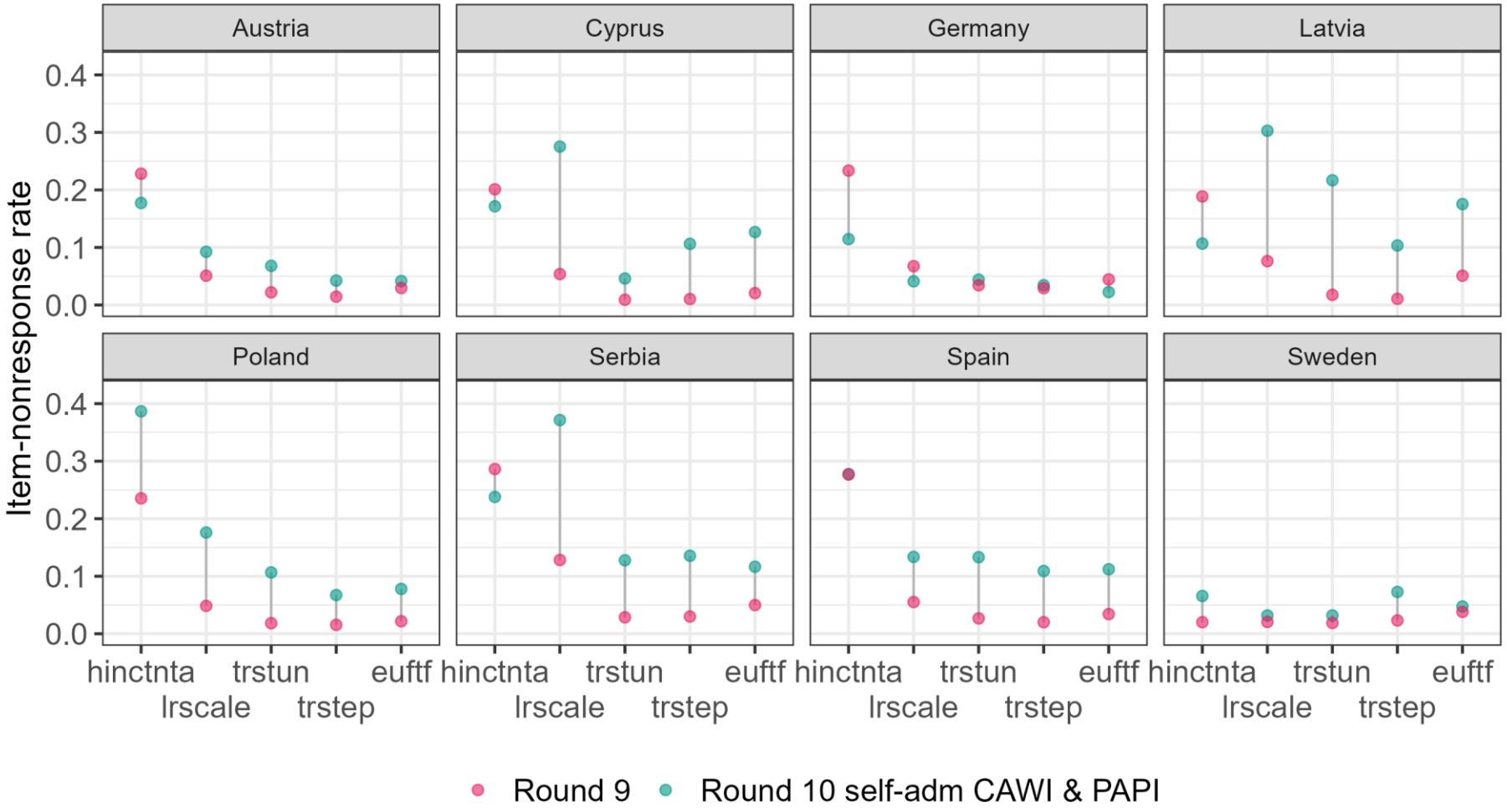
(a) more likely to be at home,

(b) more likely to respond to the interviewer.

Gender, age, education, employment, urban-rural residence, religion, immigrant status, minority status, political engagement, personality, …

Differences across countries / cultures

Mode effects



Piotr Jabkowski, Piotr Cichocki, ESS Conference 2024, Lisbon: "Does the switch to self-completion protocols deteriorate the quality of the ESS results obtained during the COVID-19 Pandemic?"

Importance of survey documentation

Documents accompanying the survey data files that describe the survey proces, including:

source questionnaires, codebooks, study descriptions, technical reports.

Extensive literature on how different elements of the survey process are linked to survey quality.

Survery documentation is essential to evaluate the quality of the survey process.

European projects: survey metadata

Dataset *Sampling and Fieldwork Practices in Europe* by Piotr Jabkowski.

Methodological information about sample types, sample design, fieldwork length, outcome rates, fieldwork control,

For six cross-national survey projects:

Eurobarometer (autumn editions),

Candidate Countries Eurobarometer (autumn editions),

European Social Survey,

European Values Study,

European Quality of Life Survey,

International Social Survey Programme (only Europe).

https://doi.org/10.5964/meth.2795

Information types

Sampling: target population definition, sampling frame, type of sample, within-HH selection of respondents

Sample design: stratification, clustering

Fieldwork: survey mode, substitution, control measures

Outcome rates: response rate or information necessary to calculate it

Table 2

Comparison of the Quality of Methodological Documentation of the Cross-Country Projects

Project abbreviation	Number of national surveys	Survey documentation quality[a]				Overall description quality[a]
		Sampling	Sample design	Fieldwork	Outcomes	
CCEB & EB (2001-2003)	84	0.750 (0.00)	0.800 (0.00)	0.333 (0.00)	0.000 (0.00)	0.471 (0.00)
EB (2004-2017)	439	1.000 (0.00)	0.800 (0.00)	0.333 (0.00)	0.000 (0.00)	0.533 (0.00)
EQLS	125	0.938 (0.11)	0.800 (0.00)	0.796 (0.07)	0.860 (0.23)	0.848 (0.05)
ESS	199	0.999 (0.02)	0.924 (0.12)	0.995 (0.03)	1.000 (0.00)	0.979 (0.03)
EVS	112	0.848 (0.21)	0.336 (0.29)	0.557 (0.36)	0.586 (0.46)	0.581 (0.28)
ISSP	578	0.841 (0.31)	0.303 (0.20)	0.710 (0.33)	0.566 (0.36)	0.605 (0.25)

Note. CCEB = Candidate Countries Eurobarometer; EB = Eurobarometer; EQLS = European Quality of Life Survey; ESS = European Social Survey; EVS = European Values Study; ISSP = International Social Survey Programme.

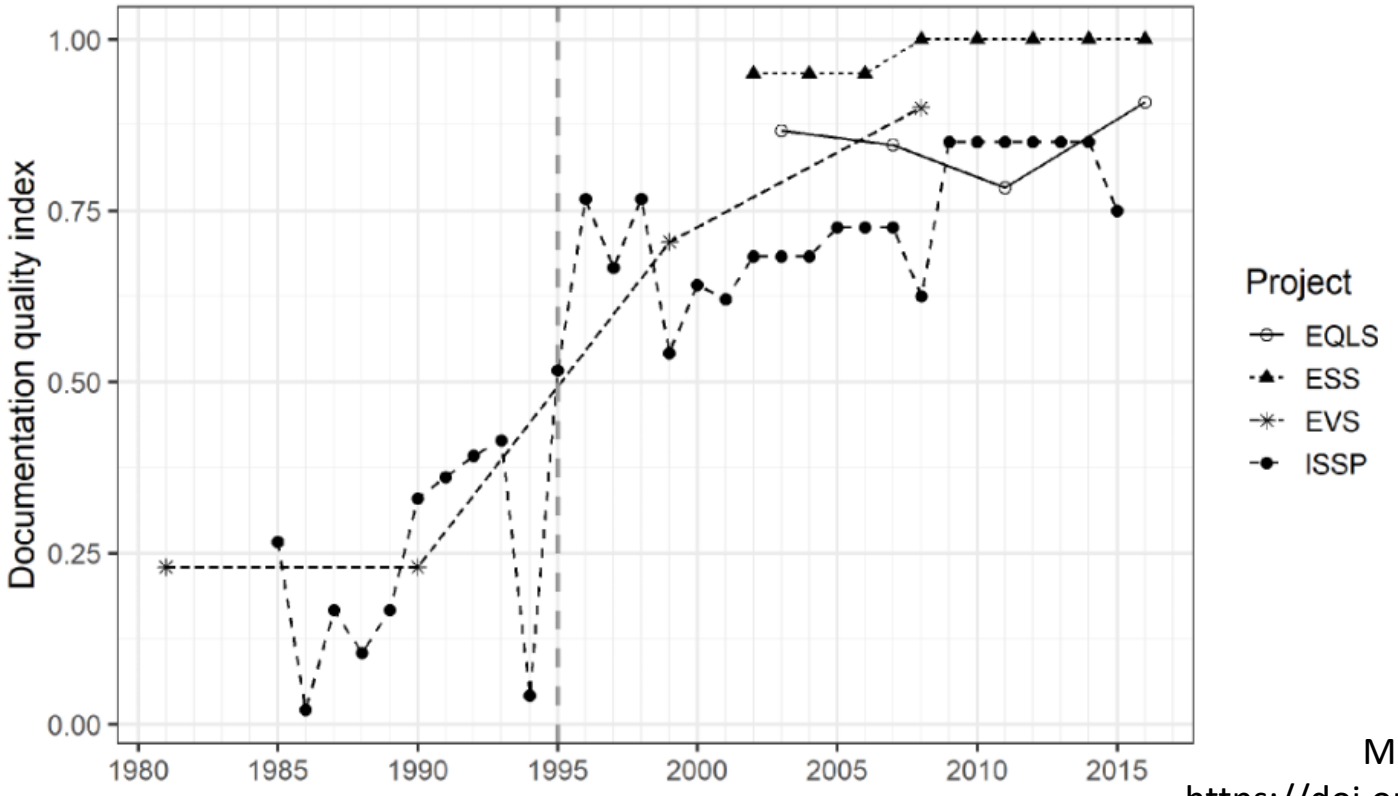
[a]Mean value of national indicators (*project*edition*country*). Values in parentheses represent standard deviations (*SD*).

Methodology, https://doi.org/10.5964/meth.2795

Documentation quality index: information on sampling, sample design, fieldwork procedures, and outcome rates.

Figure 1

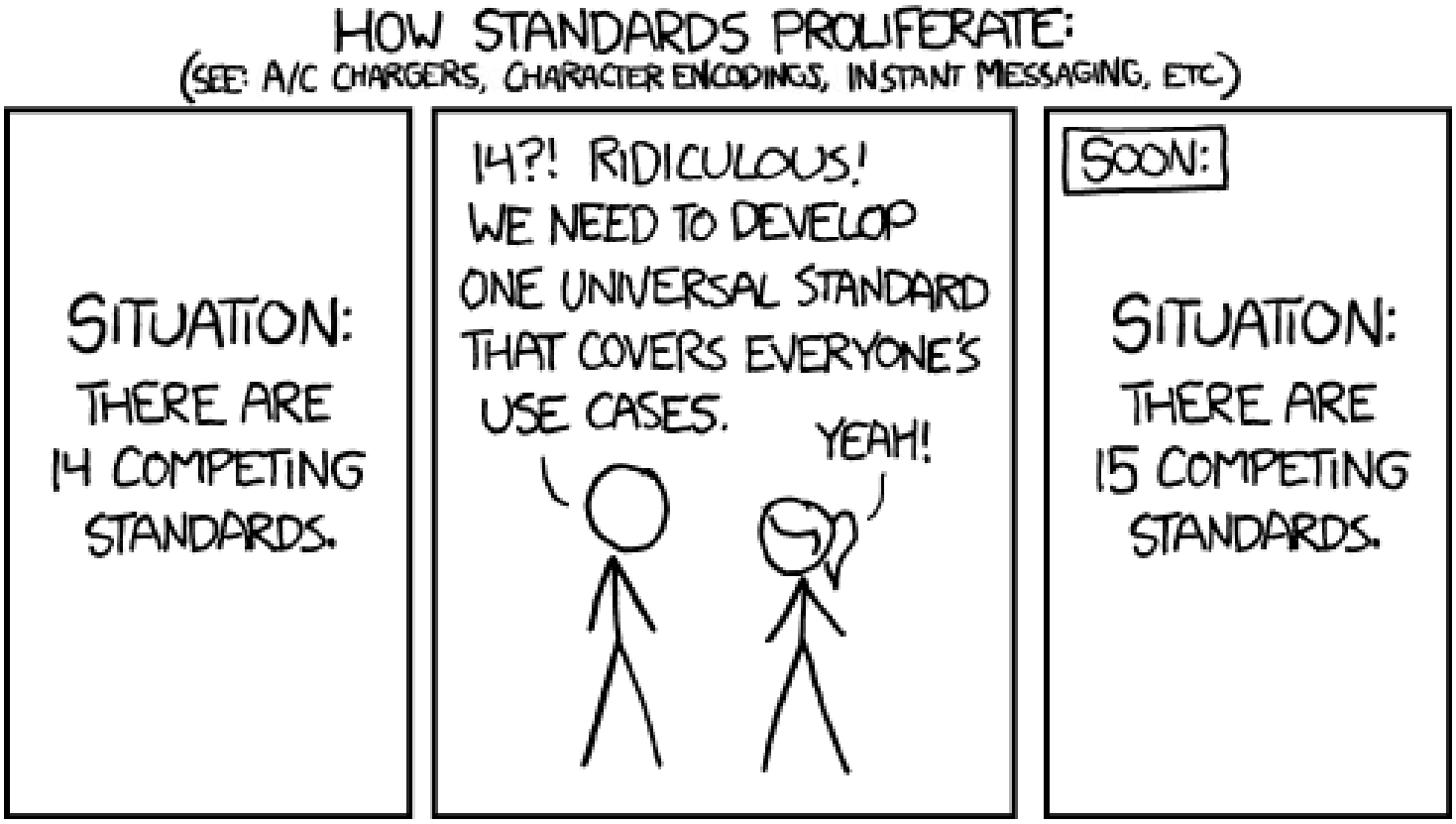
Differences in the Overall Documentation Quality in Cross-Country Projects Over Time



Methodology, https://doi.org/10.5964/meth.2795

What happened in 1995? Park, A., & Jowell, R. (1997). Consistencies and differences in a cross-national survey: The International Social Survey Programme (1995).

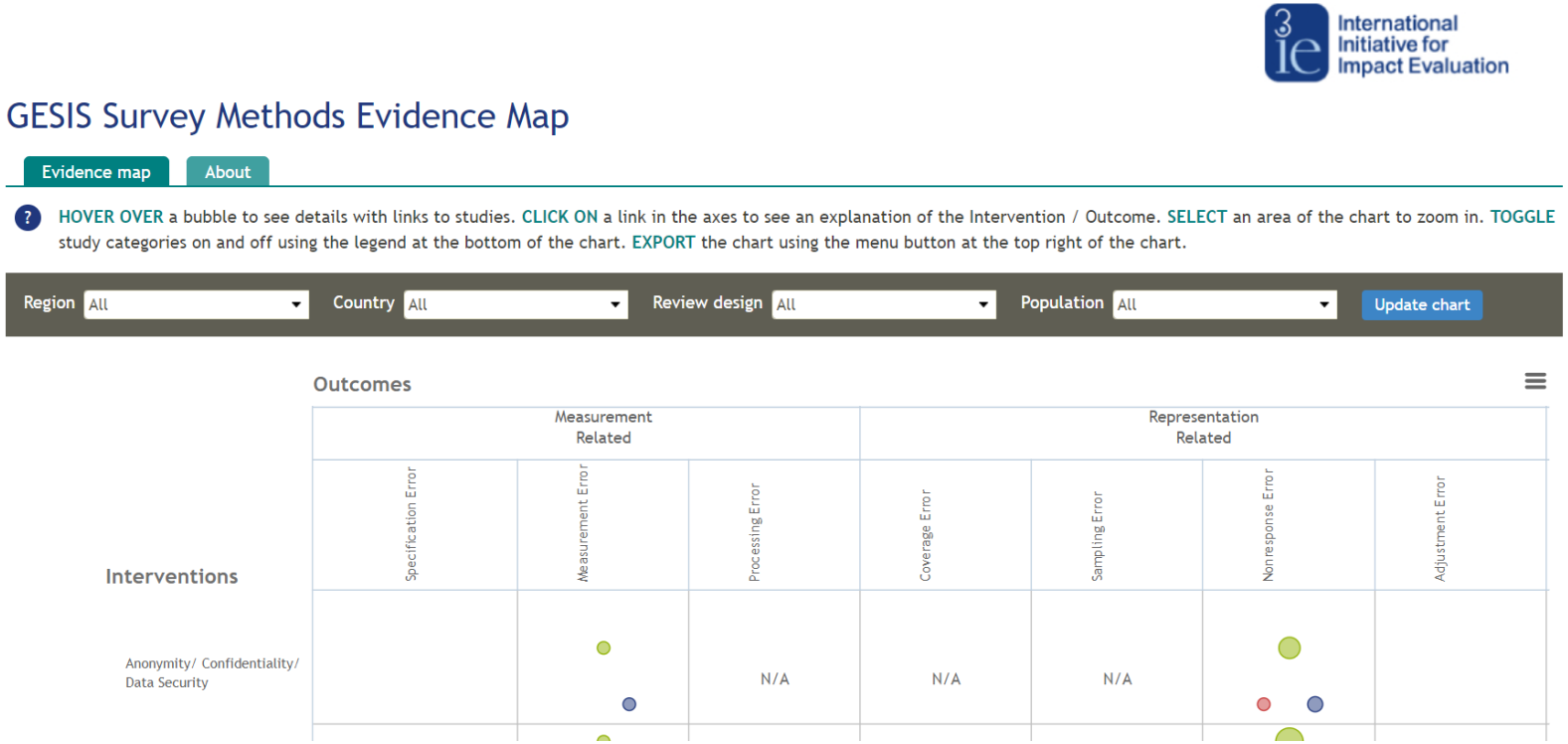
Documentation standards



https://xkcd.com/927/

GESIS Survey Methods Evidence Map

Overview of which aspects of the survey proces have an affect on what aspect of total survey error: https://egmopenaccess.3ieimpact.org/evidence-maps/gesis-survey-methods-evidence-map



Survey Quality Predictor

Links formal and linguistic characteristics of survey questions to measurement quality based on Multi-Trait Multi-Method (MTMM) experiments in the European Social Survey and other projects.

https://sqp.gesis.org/

Moved to Gesis, Germany, from UPF.

SQP 3.0 Survey Quality Predictor

Database

Fixable problems

- Some differences seem to be inconsequential, e.g. the difference between „trust” and „confidence” in items on trust in institutions.
- Some issues can be corrected, e.g. some deviations from sample representativeness.
- Some issues are disqualifying, e.g. omitting large parts of a country’s territory or lack of representation of important population groups; errors in translation.
- No level of statistical expertise will help if the data are very bad.

Problems of scale and size

Problem of scale and data subsets. Difference between 1 strongly biased survey among 200 versus 1 strongly biased survey in a subset of 5 surveys from Albania.

To sum up

Data quality is at the core of each comparative analysis.

Quality screening is a prerequisite for the application of statistical procedures, including harmonization.

Only surveys that meet some minimum quality criteria can be analyzed together.