GESIS Leibniz Institute
for the Social Sciences

Evaluating Total Survey Quality:

The Case of Surveys Conducted During the COVID-19 Pandemic in Germany

Leibniz Association

Karolina von Glasenapp, Thomas Skora, Tobias Gummer, Elias Naumann
GESIS – Leibniz Institute for the Social Sciences, Mannheim Germany

Project: Survey Data Collection and the COVID-19 Pandemic (SDCCP)

SPONSORED BY THE

Federal Ministry
of Education
and Research

International Total Survey Error Workshop, 18th September 2024
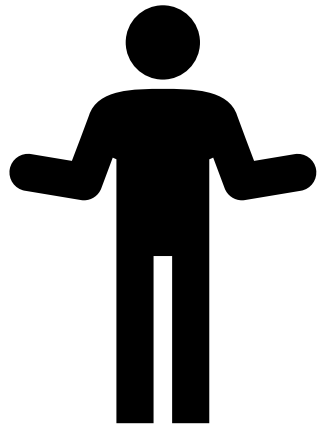
Survey quality in general

- A *multidimensional* concept -> importance of a *comprehensive* empirical assessment

- However, practical constraints due to high demands on resources and information available

- Therefore, large-scale assessments of the survey landscape are lacking so far
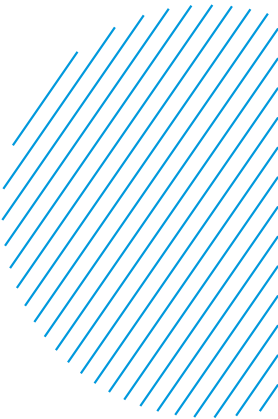
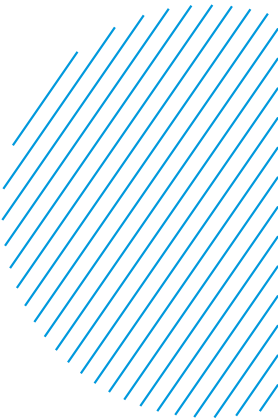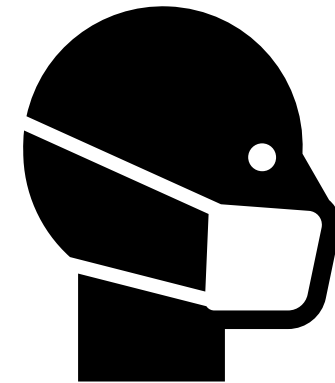# Survey quality in times of crises
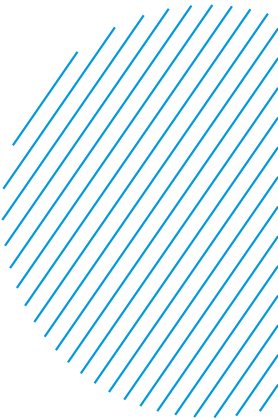
Reliable information

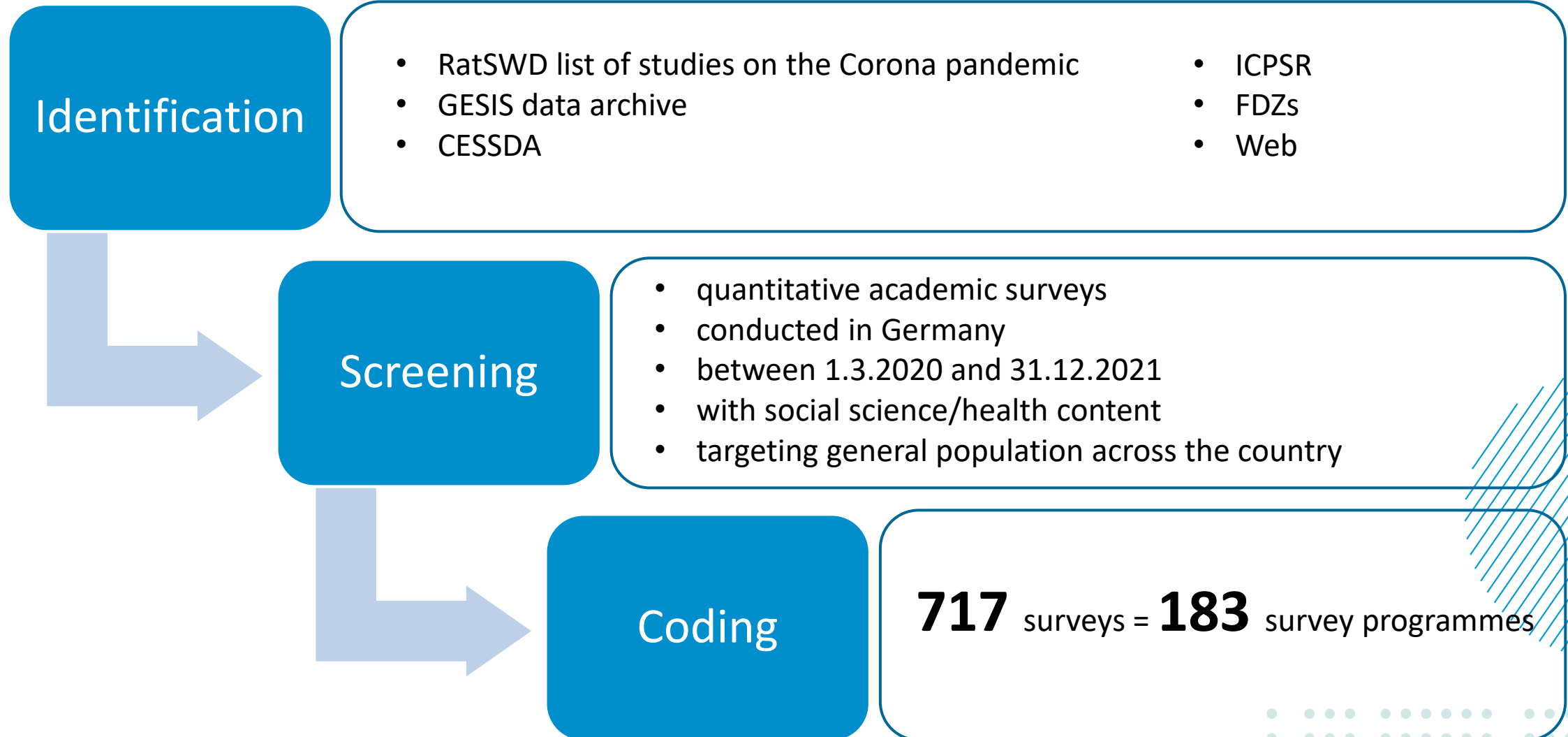Readily available

Methodological cha(lle)nges

Research questions

RQ1. What was the quality of social science surveys conducted during the COVID-19 pandemic in Germany?

RQ2. How were different dimensions of survey quality associated with each other during the COVID-19 pandemic?

GESIS Leibniz Institute for the Social Sciences

Data selection process

Identification
- RatSWD list of studies on the Corona pandemic
- GESIS data archive
- CESSDA
- ICPSR
- FDZs
- Web

Screening
- quantitative academic surveys
- conducted in Germany
- between 1.3.2020 and 31.12.2021
- with social science/health content
- targeting general population across the country

Coding
717 surveys = **183** survey programmes

GESIS Leibniz Institute
for the Social Sciences

Method – Theoretical background

Data producers' perspective

&

Data users' perspective

Total Survey Error framework

Groves et al. (2009),
Lyberg and Weisberg (2016)

Total Survey Quality framework

Biemer (2010)

FAIR Principles

Wilkinson et al. (2016),
Eder and Jedinger (2019)

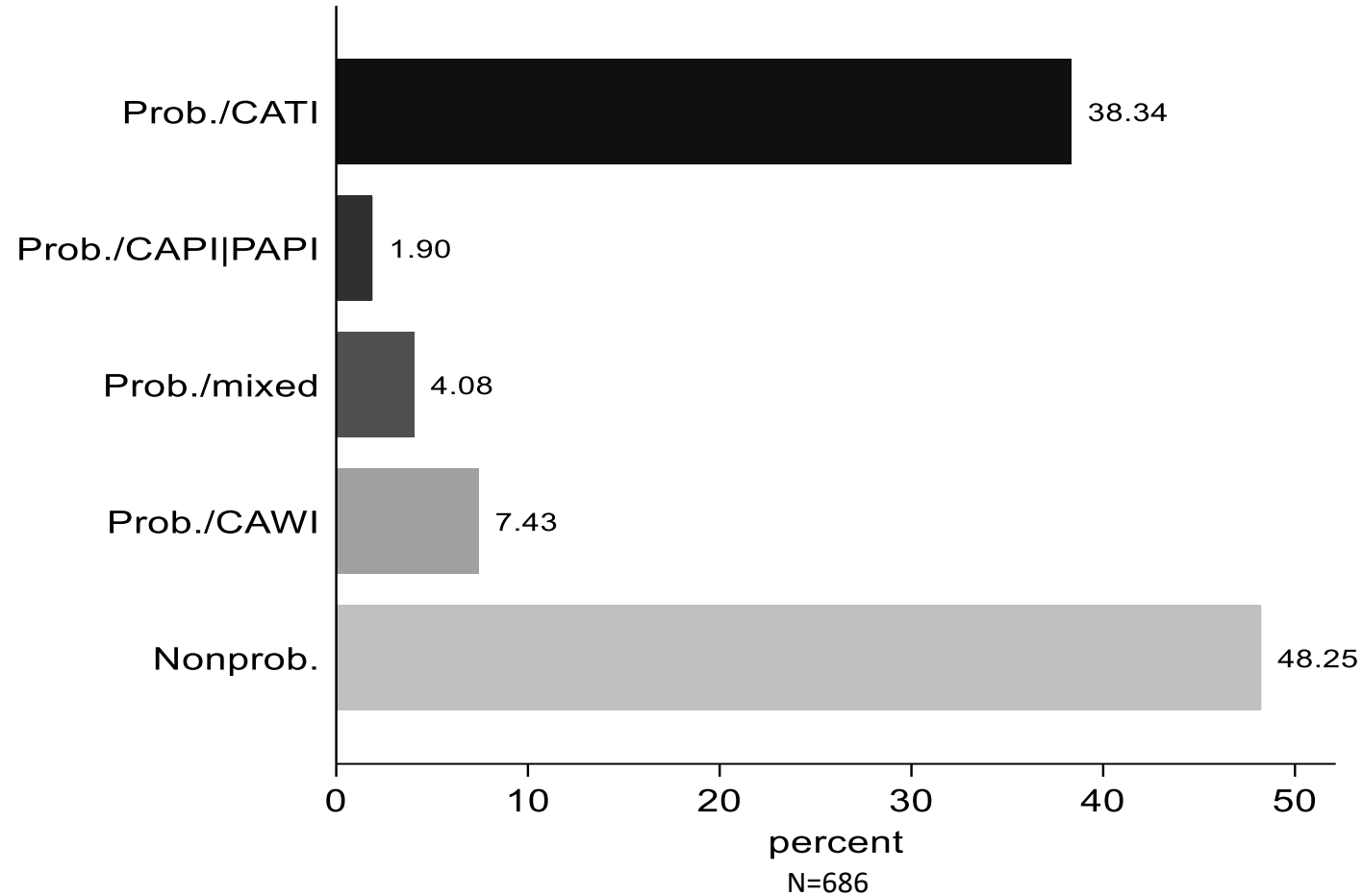GESIS Leibniz Institute for the Social Sciences

Method – Applied framework

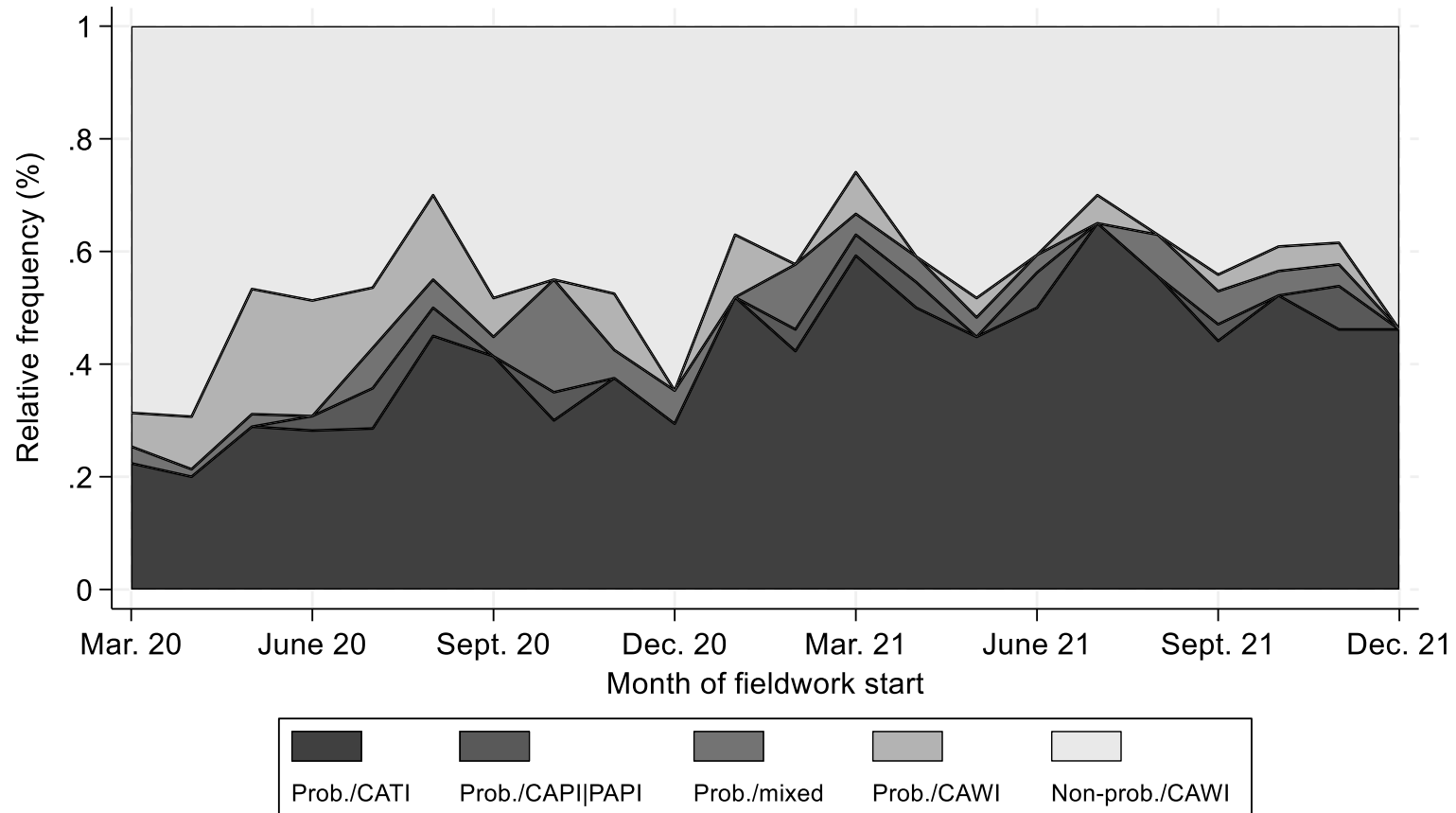| Perspective	Quality dimension	Definition
Data producer	Accuracy	Total Survey Error is minimized
Data user	Interpretability	Information on survey design is available
Data user	Accessibility	Access to the first results and data is provided quickly

Method – Applied framework

Quality dimension	Operationalization
Accuracy	3 descending categories (based on empirical literature): 1. Prob. & other-than-only CAWI 2. Prob. & CAWI only 3. Nonprob. & CAWI
Interpretability	Additive index based on the 0/1 coding of the following variables: • Target population • Concrete sampling procedure • Sample size • Date of data collection (at day level) • Any outcome rate Index range between 0 (minimum) and 5 (maximum)
Accessibility	Time gap between the beginning of fieldwork and the publication of the first results and data (monthly basis)

Accuracy – Distribution of survey designs

Prob./CATI — 38.34
Prob./CAPI|PAPI — 1.90
Prob./mixed — 4.08
Prob./CAWI — 7.43
Nonprob. — 48.25

percent
N=686

Accuracy – Development over time

N=686

Interpretability

M = 3.78
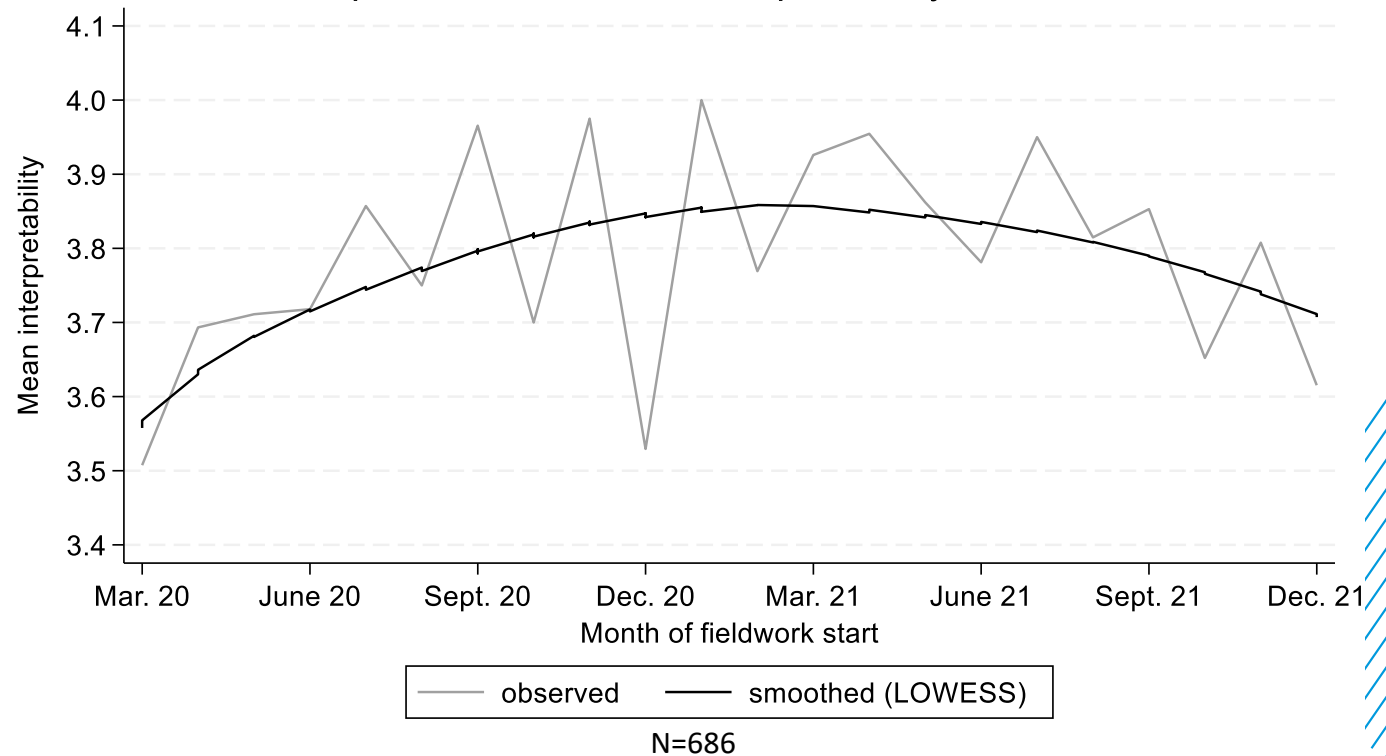
67% of surveys scored 4+ (max. 5)

„Any outcome rate" as the worst-performing component

Development of the mean interpretability index over time

N=686

Accessibility of first results

51.6% of surveys – publication within 2 months after fieldwork start

19.2% of surveys – no publication observed

Development of the proportion of surveys with results publication within 2 months

N=483
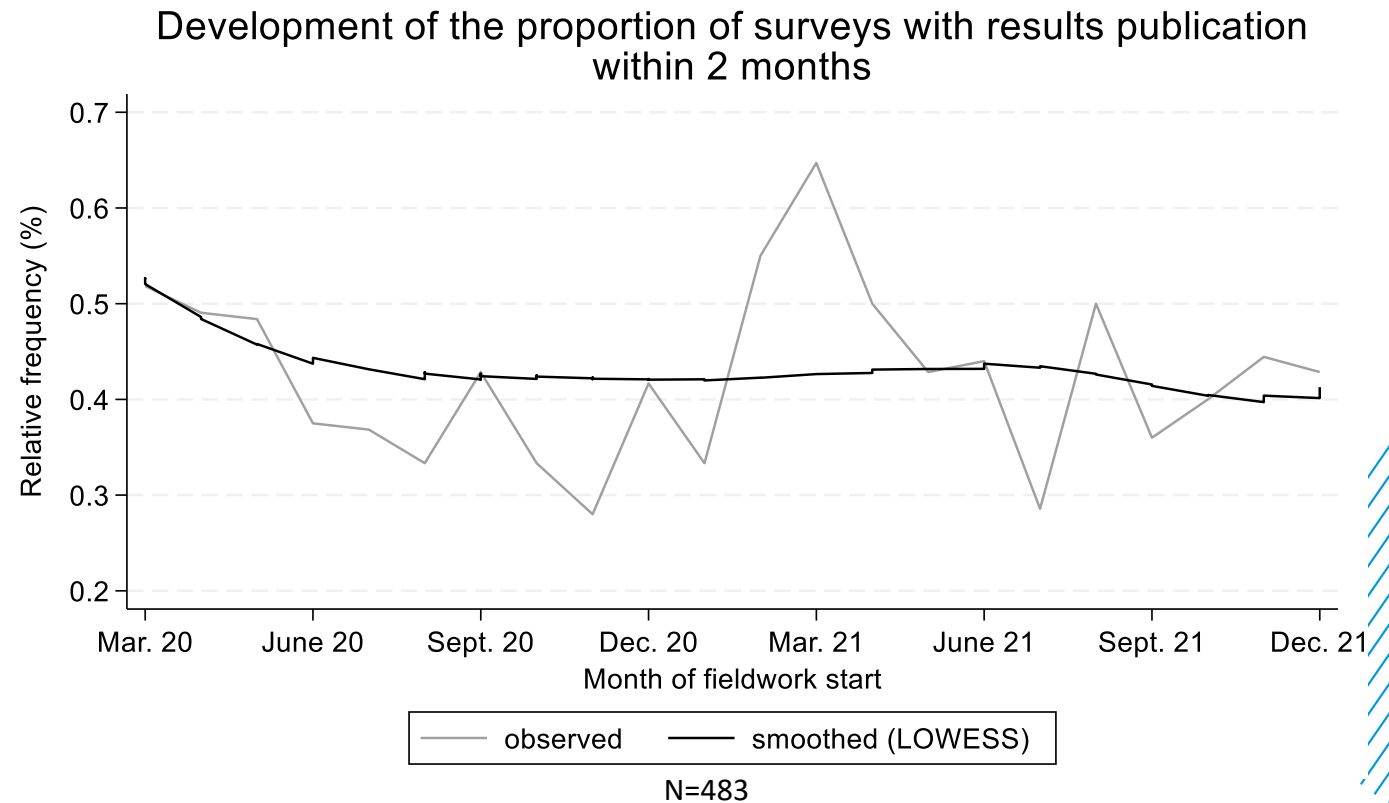
Accessibility of data

14.2% of surveys – publication within 2 months after fieldwork start, 37.5% within 12 months

36.3% of surveys – no publication observed

Development of the proportion of surveys with data publication within 12 months

N=590
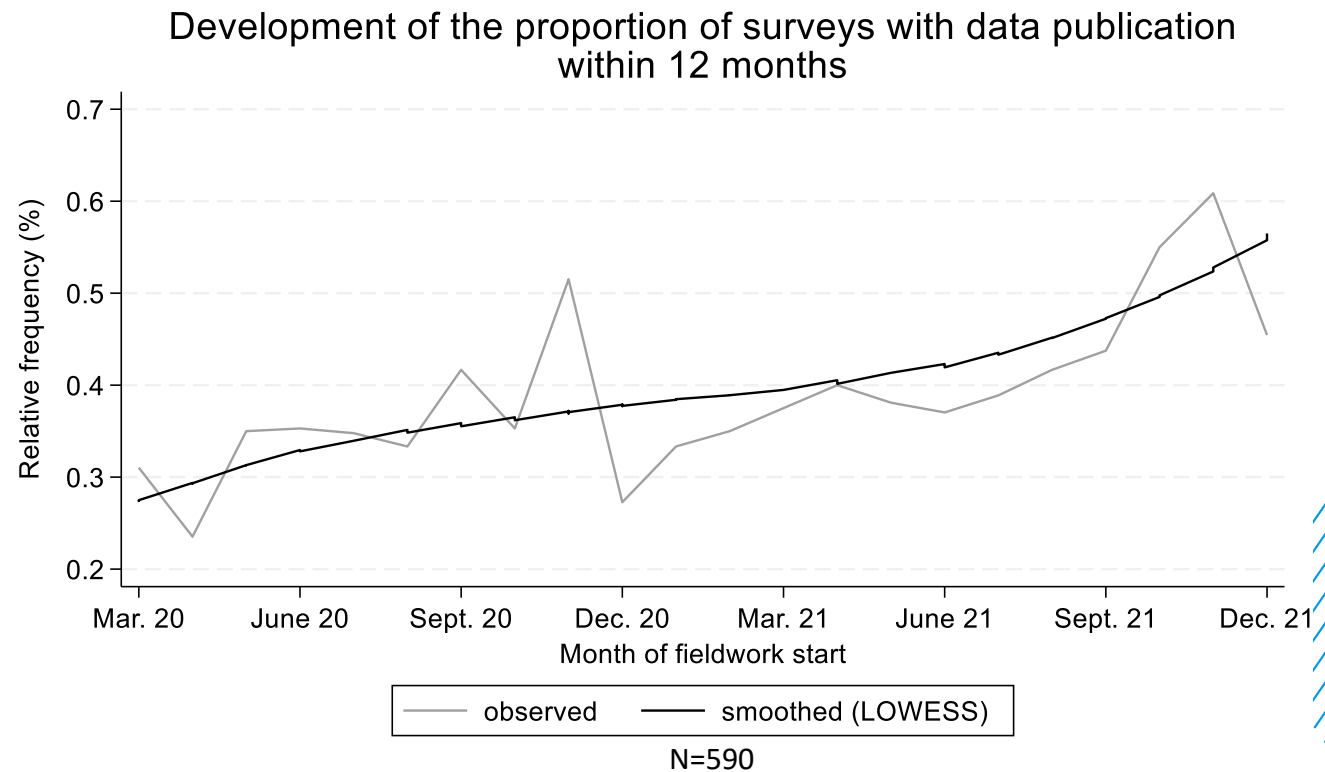
Accuracy and interpretability

	Model 1	Model 2
(Fieldwork start)	3.590**	2.716**
	(1.478)	(1.362)
(Fieldwork start)^2	-0.002**	-0.002**
	(0.001)	(0.001)
Survey design (Ref. Non-probability)		
Prob. CATI	--	0.240
		(0.238)
Prob. CAWI	--	1.195***
		(0.158)
Prob. mixed mode	--	0.672***
		(0.239)
Prob. CAPI or PAPI	--	1.028***
		(0.269)
Constant	-1,312.322**	-993.428**
	(541.986)	(499.506)
N surveys	686	686
N observations	686	686
Adj. R^2	0.014	0.184

Note: OLS regression, *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Probability surveys were more likely to provide relevant information than non-probability surveys

Accuracy and accessibility of first results

	Model 1	Model 2
(Fieldwork start)	0.026	-0.006
	(0.021)	(0.022)
(Fieldwork start)^2	0.004*	0.001
	(0.002)	(0.003)
(Fieldwork start)^3	-0.000	-0.000
	(0.000)	(0.000)
Survey design (Ref. Non-probability)		
Prob. CATI	--	0.582***
		(0.166)
Prob. CAWI	--	-0.868**
		(0.225)
Prob. mixed mode	--	-1.479***
		(0.284)
Prob. CAPI or PAPI	--	-0.220
		(0.265)
Constant	-1.125***	-1.059**
	(0.128)	(0.146)
N surveys	483	483
N observations	4,613	4,613
Pseudo R^2	0.142	0.171

Note: event history analysis, *** p<0.01, ** p<0.05, * p<0.1

Non-probability CAWI surveys were associated with an increased likelihood to publish results compared to prob. CAWI and mixed mode surveys, but not compared to prob. CATI surveys

Accuracy and accessibility of data

	Model 1	Model 2	
(Fieldwork start)	0.038***	0.019	
	(0.011)	(0.012)	
(Fieldwork start)^2	-0.002	0.000	
	(0.002)	(0.002)	
Survey design (Ref. Non-probability)			
Prob. CATI	--	0.905***	Probability surveys were associated with an increased likelihood to publish data compared to non-prob. surveys
		(0.146)	
Prob. CAWI	--	1.547***	
		(0.198)	
Prob. mixed mode	--	1.080***	
		(0.226)	
Prob. CAPI or PAPI	--	1.345***	
		(0.315)	
Constant	-3.359**	-3.610***	
	(0.145)	(0.155)	
N surveys	590	590	
N observations	11,740	11,740	
Pseudo R^2	0.076	0.101	

Note: event history analysis, *** p<0.01, ** p<0.05, * p<0.1

GESIS Leibniz Institute
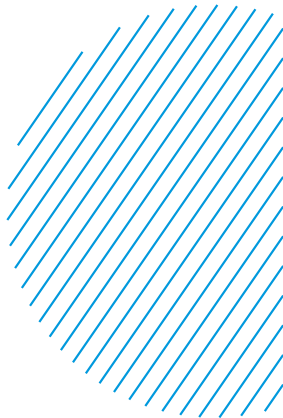for the Social Sciences

RQ1 – Survey quality during the pandemic

At the beginning of the pandemic:
- prevalence of surveys of lower accuracy
- focus on the publication of results

Later in the course of the pandemic:
- increase in accuracy and interpretability
- a stronger focus on the publication of data

RQ2 – Associations between quality dimensions

A trade-off between quality dimensions of accuracy and accessibility of results: surveys of lower accuracy related to quick publication of results

On the contrary, surveys of higher accuracy associated with higher interpretability and accessibility of data
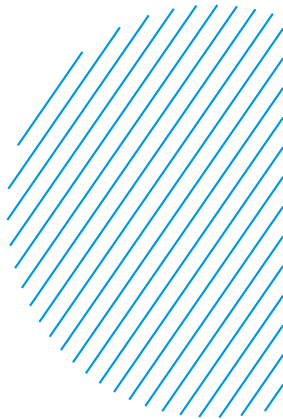
References

Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, *74*(5), 817–848. https://doi.org/10.1093/poq/nfq058

Eder, C., & Jedinger, A. (2019). FAIR national election studies: How well are we doing? *European Political Science*, *18*(4), 651–668. https://doi.org/10.1057/s41304-018-0194-3

Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). John Wiley & Sons.

Lyberg, L. E., & Weisberg, H. F. (2016). Total survey error: A paradigm for survey methodology. In C. Wolf, D. Joye, T. W. Smith, & Y. Fu (Eds.), *The SAGE Handbook of Survey Methodology*. SAGE.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), Article 1. https://doi.org/10.1038/sdata.2016.18
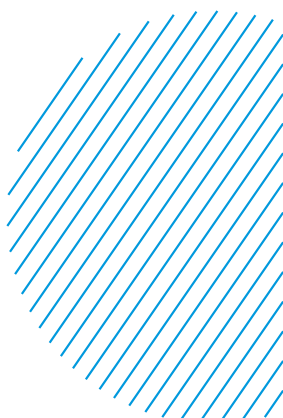
Data sets available for download

Data sets generated within the project are freely available for download at GESIS archive:

SDCCP 1 - Survey Design and Quality During the Covid-19 Pandemic

https://doi.org/10.7802/2652

SDCCP 2 - Survey Design of Longitudinal Surveys Before and During the Covid-19 Pandemic

https://doi.org/10.7802/2759

Thank you for your attention

Karolina von Glasenapp

GESIS – Leibniz Institute for the Social
Sciences
B6, 4-5 | 68519 Mannheim, Germany

Karolina.Glasenapp@gesis.org

Data set - variables

Basic information on the survey	• e.g., survey initiator and survey conductor, fieldwork dates, funding
Survey design characteristics	• e.g., mode, sampling procedure, sample size, use of incentives and reminders
Availability	• of survey data, survey documentation and the first results

Interrater reliability – categorical variables

	Total	Round 1	Round 2
Percent agreement	0.92	0.90	0.93
Brennan and Prediger's coefficient	0.87	0.85	0.90

Note: N(total) = 20, N(round) = 10

Interrater reliability – continuous variables

	Total	Round 1	Round 2
Percent agreement	0.90	0.87	0.93
Interclass correlation coefficient	0.99	0.99	0.94

Note: N(total) = 20, N(round) = 10