Survey design features that matter: A meta-analysis using official statistics surveys of the Netherlands

Jeldrik Bakker, Jonas Klingwort, Vera Toepoel

International Blaise Users Conference 2025, Rotterdam, 7 – 11 april 2025
08.04.2025

Introduction

- Objective: Study the impact of survey design choices on response rates.
- Context: Growing importance of reliable survey data for policy and research.
- Approach: Meta-analysis of official statistics surveys conducted in the Netherlands.
- Key Question: What design elements, known before and after data collection, influence survey response rates?

Background

Studies show that survey design features, such as ...

- incentives [1],
- data collection modes [2, 3, 4, 5, 6],
- recruitment strategy [7],
- fieldwork duration and strategy [8],
- questionnaire length [9],
- sponsorship [10], and
- (perceived) response burden [10],

... have effects on the response rates.

Previous study by [11], using the example of crime surveys in Germany.

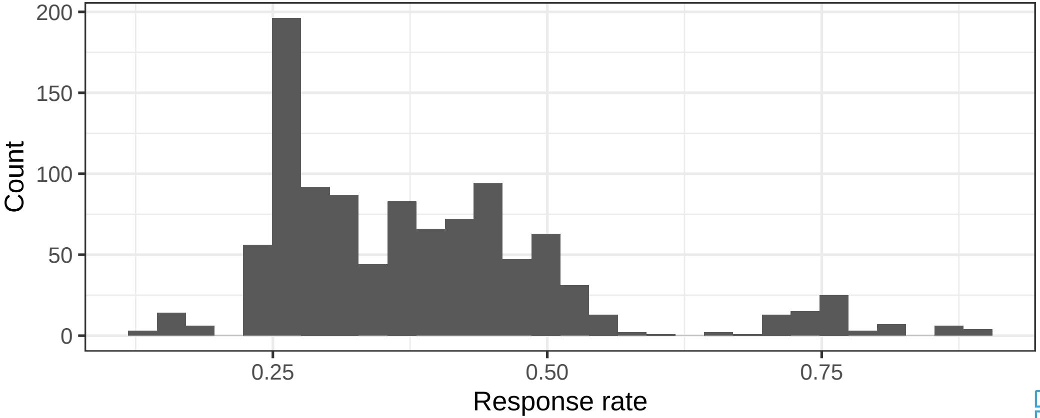
Data (Meta-analysis)

- 45 surveys (Level 3), with 105 editions (Level 2), and 1046 samples (Level 1)
- 2018 – 2024
- Individual and household surveys
- Min. sample size = 275, response size = 100
- Max. sample size = 56269, response size = 21610
- Total sample size: 7 mio., total response size = 2.5 mio.
- 61 design features: survey, questionnaire, respondent [before and after data collection]

Features

- Survey [before]: year, month, fieldwork period, periodicity, data collection mode, interviewer, topic, follow-up, panel-indicator, incentive, number of letters, number of reminders, email, number of contacts
- Questionnaire [before]: min. and max. for number of blocks, pages, questions, single choice, multiple choice, numeric questions, open questions, datetime, empty
- Respondent [before]: age and sex distributions in sample (fractions)
- Questionnaire [after]: median for blocks, pages, questions, single choice, multiple choice, numeric, open, datetime, empty, answers, duration
- Respondent [after]: age and sex distribution in response; fraction for cawi, desktop, smartphones, other devices

Target



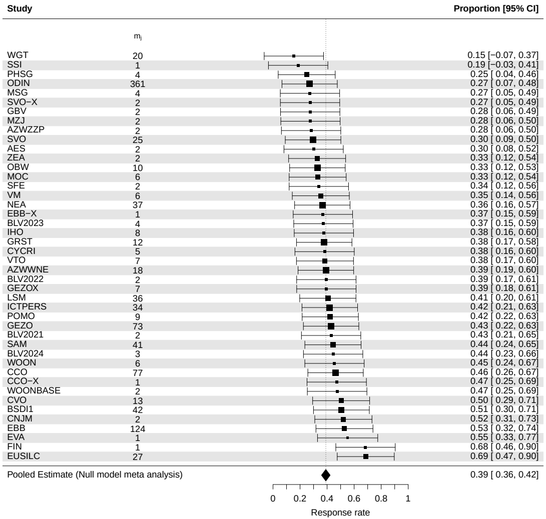
Meta-analysis and Meta-regression

- Meta analysis (test for heterogeneity)
- Multi-level meta-regression model (explain heterogeneity)
- Accounts for within- and between-study variability
- Random-effects structure to handle nested data
- Heterogeneity: I^2 and Q statistics to assess heterogeneity

Meta analysis and meta-regression models

- M0: no features
- M1: survey [b]
- M2: survey [b], questionnaire [b]
- M3: survey [b], questionnaire [b], respondent [b]
- M4: survey [b], questionnaire [b], respondent [b], questionnaire [a]
- M5: survey [b], questionnaire [b], respondent [b], questionnaire [a], respondent [a]
- M6: AIC optimized

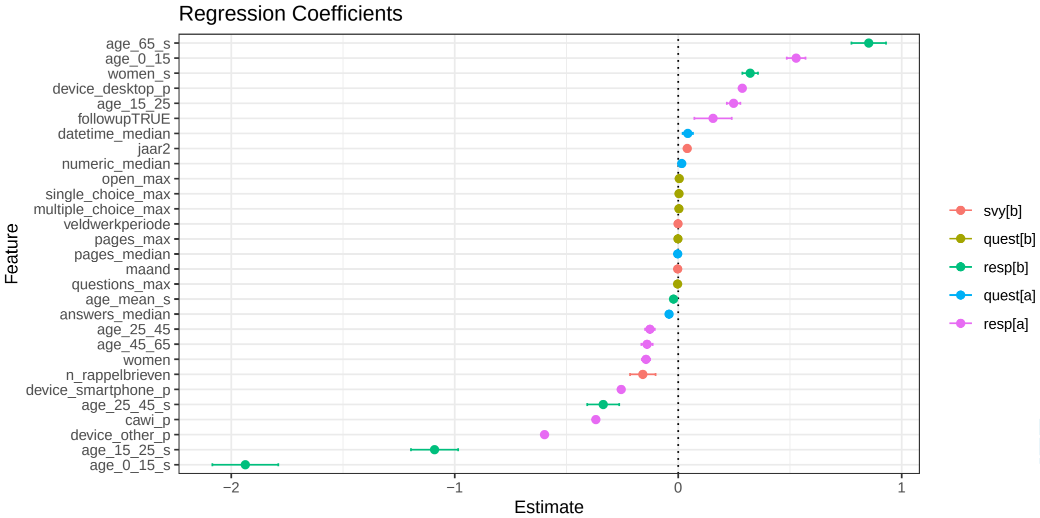
Results - Meta analysis forest plot (M0)



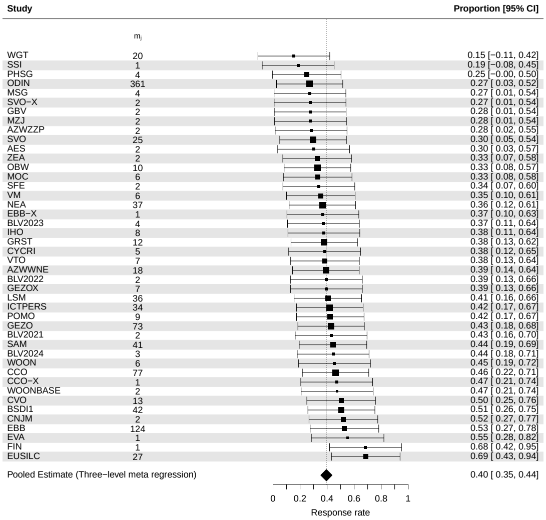
Results - Variance components

	Meta-analysis			Meta regression			
	M0	M1	M2	M3	M4	M5	M6
Level 1	0.256	0.170	0.403	0.388	0.293	0.119	0.175
Level 2	15.20	43.86	99.60	99.61	76.85	16.44	17.32
Level 3	84.55	55.97	0.000	0.001	22.86	83.44	82.51
I^2	99.74	99.83	99.60	99.61	99.71	99.88	99.83
Q	$< .0001$	$< .0001$	$< .0001$	$< .0001$	$< .0001$	$< .0001$	$< .0001$
f	0	14	32	39	50	61	30
AIC	90275	76585	76600	73378	73408	51984	51859

Results - Meta regression (M6)



Results - Meta regression forest plot (M6)



Discussion

- Predictors help reduce variance at Level 1 and Level 2 but not fully at Level 3.
- Biggest drop in AIC from M0 to M1 (adding survey[b]); M2–M4 (questionnaire[b], respondent[b], questionnaire[a]) do not improve; big drop with M5 (respondents[a]); M6 comparable to M5, but half the number of features.
- M6 achieves the lowest AIC, however, with 29 predictors, it remains relatively complex, and further simplification may still be beneficial.
- M6: strongest relationships between the moderators resp[b][a] and the response rate; weak relationships between svy[b], quest[b][a] and the response rate
  - Features where we have less control show stronger relationships
- I^2 remains extremely high across all models, confirming substantial residual heterogeneity, i.e., unexplained heterogeneity persists despite adding predictors (Q statistics confirms).
- Next steps: exploring higher-order interactions and nonlinear effects to better capture study-level variance.

Conclusion

- Comprehensive insights into understanding response rates for official statistics surveys.
- Potential to support effective data collection strategies and survey designs.
- Practical toolbox for national statical agencies, survey agencies, and survey researchers.
- First study on this scale in official statistics and first application of the Framework of Survey Behavior.

References I

[1] Vera Toepoel. Effects of incentives in surveys. In *Handbook of Survey Methodology for the Social Sciences*, pages 209–223. Springer, New York, USA, 2012.

[2] Corey Greenlaw and Sharon Brown-Welty. A Comparison of Web-Based and Paper-Based Survey Methods: Testing Assumptions of Survey Mode and Response Cost. *Eval. Rev.*, 33(5):464–480, 2009.

[3] Don A. Dillman, Glenn Phelps, Robert Tortora, Karen Swift, Julie Kohrell, Jodi Berck, and Benjamin L. Messer. Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Soc. Sci. Res.*, 38(1):1–18, 2009.

References II

[4] Cauane Blumenberg and Aluísio J. D. Barros. Response rate differences between web and alternative data collection methods for public health research: a systematic review of the literature. *Int. J. Public Health*, 63(6):765–773, 2018.

[5] Maria Krysan, Howard Schuman, Lesli Jo Scott, and Paul Beatty. Response rates and response content in mail versus face-to-face surveys. *Public Opin. Q.*, 58(3):381–399, 1994.

[6] Margaret Farnworth, Katherine Bennett, and Vincent M. West. Mail vs. telephone surveys of criminal justice attitudes: A comparative analysis. *J. Quant. Criminol.*, 12(1):113–133, 1996.

[7] Martha Sinclair, Joanne O'Toole, Manori Malawaraarachchi, and Karin Leder. Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches. *BMC Med. Res. Method.*, 12(1):1–8, 2012.

References III

[8] Jonas Klingwort, Bart Buelens, and Rainer Schnell. Early versus late respondents in web surveys: Evidence from a national health survey. *Statistical Journal of the IAOS*, 34(3):461–471, 2018.

[9] Mirta Galesic and Michael Bosnjak. Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opin. Q.*, 73(2):349–360, 2009.

[10] Don A. Dillman. Towards survey response rate theories that no longer pass each other like strangers in the night. In Philip S. Brenner, editor, *Understanding survey methodology: Sociological theory and applications*, pages 15–44. Springer, Switzerland, 2020.

[11] Jonas Klingwort and Vera Toepoel. Effects of survey design features on respons rates: a meta-analytical approach using the example of crime surveys. *arXiv*, 202