

Effects of Survey Design Features on Response Rates: A Meta-Analytical Approach Using the Example of Crime Surveys

Jonas Klingwort¹  · Vera Toepoel¹

¹Statistics Netherlands, Research & Development

When conducting a survey, many choices regarding survey design features have to be made. These choices affect the response rate of a survey. This paper analyzes the individual effects of these survey design features on the response rate. For this purpose, data from a systematic review of crime surveys conducted in Germany between 2001–2021 were used. First, a meta-analysis of proportions is used to estimate the summary response rate. Second, a meta-regression was fitted, modeling the relationship between the observed response rates and survey-design features, such as the study year, target population, coverage area, data collection mode, and institute. The developed model informs about the influence of certain survey design features and can predict the expected response rate when (re-) designing a survey. This study highlights that a thoughtful survey design and professional survey administration can result in high response rates.

Keywords: meta-analysis; meta regression; data collection mode; fieldwork; non-response; victimization; data quality

1 Introduction

The achieved response rate of a survey depends on several design features, for example, the survey's length and complexity, offered incentives, or the reputation and credibility of the conducting survey agency (Dillman 2020). In recent years, declining trends in survey response rates have been observed (Meyer et al. 2015). Nevertheless, these findings might not be generalized as well-designed surveys can still yield high response rates (Holtom et al. 2022). Consequently, it is important to know in which direction specific survey design features influence the response rate. With this

knowledge, tailored and specific survey design features can be chosen when conducting a population survey.

This work aims to quantify the effects of survey design features on the response rate. This approach is demonstrated using German crime surveys. No regular or repeated standardized official crime survey is conducted in Germany. Instead, criminological survey research in Germany is dominated by individual and independent surveys. These differ, for example, by contractor, executive survey agency, target population, sampling design, or data collection mode. This circumstance allows to study the effects of different survey design features and their impact on the response rate.

For this study, the initial systematic review by Klingwort (2017) is used but has been updated. The updated systematic review considers German crime surveys between 2001–2021 and collected survey design feature information based on survey-related publications. The survey design features, such as study year, target population, coverage, data collection mode, responsible institute/contractor, sample, and response size, were extracted from these publications.

First, a meta-analysis of proportions is used to estimate the summary effect size, i.e., the response rate. Second, using a meta-regression, the effects of the survey design

Supplementary Information The online version of this article (<https://doi.org/10.18148/srm/8173>) contains supplementary material.

Corresponding author: Jonas Klingwort, Research & Development, Statistics Netherlands, CBS-weg 11, 412 EX Heerlen, The Netherlands (Email: j.klingwort@cbs.nl)

features on the response rate are estimated. Moreover, it will be informed on the models' prediction quality, optimal and non-optimal survey design feature sets, and potential model selection.

2 Background

Response rates of population surveys show a decline during the last decades (Meyer et al. 2015; Czajka and Beyler 2016; Williams and Brick 2018; Luiten et al. 2020; Daikeler et al. 2020; Dutwin and Buskirk 2021; Lugtig et al. 2022). For crime surveys even lower response rates can be expected since sensitive topics are surveyed (Tourangeau and Smith 1996). Nevertheless, the number of conducted surveys is enormous and increasing (Singer 2016). For example, Presser and McCulloch (2011) reported that the number of surveys conducted between 1984–2004 increased at a rate many times greater than the population in the US. However, in the context of decreasing survey response and increasing costs for data collection, the future of surveys has been questioned (Couper 2013, 2017; Alwin 2013; Rao and Fuller 2017), and as a result, attempts and initiatives emerged in recent years to find alternatives to surveys (Link et al. 2014; Galesic et al. 2021). From these developments, we derive the importance of this study. There is an urgent need to ensure that response rates do not decline further and, ideally, start to increase again. In an ideal situation, a survey would be designed so that no missing data would occur (Allison 2002). However, this ambitious goal can, at best, be approached. A well-designed survey and professional administration are essential in approaching this solution. In this work, we will provide answers and recommendations regarding this solution.

Dillman (2020) identified seven factors related to survey design and administration that influence survey response rates. First, the choice of data collection mode or combination of data collection modes affects the response rate. For example, CATI surveys achieve only low response rates (Kohut et al. 2012; Habermann et al. 2017; Olson 2020). In recent years, web surveys have become more prominent in survey research. However, web surveys also commonly achieve lower response rates than other modes (Blumenberg and Barros 2018; Daikeler et al. 2019). Due to the particular advantages and disadvantages of different modes, their combinations, and advancing technical development, there have been developments in adaptive and responsive survey designs (Groves and Heeringa 2006; Axinn et al. 2011; Schouten et al. 2013; Luiten and Schouten 2013; Laflamme and Wagner 2016; Chun et al. 2018), mixed-mode or multiple-mode surveys (de Leeuw 2005; Greenlaw and Brown-Welty 2009; Dillman et al. 2009; Millar and Dillmann 2011; Dillmann et al. 2014). In an adaptive survey design, e.g.,

the data collection mode or contact times are adjusted during the fieldwork. Surveys using a responsive design have sequential phases, wherein in each phase, different fieldwork strategies are applied (Chun et al. 2018). Mixed-mode surveys use different data collection modes for different subgroups, and the respondents in multiple-mode surveys are interviewed at one time with different modes (Schnell 2019).

Second, the sponsorship of the survey affects the response rate. For example, there is evidence that surveys of governmental bodies and agencies achieve higher response rates than commercial survey agencies (Brick and Williams 2012). Therefore, a negative connotation about the sponsor will result in substantive adverse effects on the response rate (Faria and Dickinson 1996; Edwards et al. 2014; Isani and Schlipphak 2022).

Third, the (perceived) response burden of the survey affects the response rate. Here, factors such as the questionnaire length (Galesic and Bosnjak 2009), motivation or perception of the survey (Yan et al. 2020), and the respondent's satisficing strategies (Krosnick et al. 1996) will affect the response rate. A recent review and conceptual framework on factors contributing to the (perceived) response burden is given by Yan and Williams (2022).

Fourth, the offered (appropriate) incentives affect the response rate. There is a large body of literature showing that incentives increase response. However, monetary incentives are more effective than gifts. In addition, incentives received in advance work better than those received after survey participation. However, the optimal size of incentives is debated, so the use of tailored incentives is recommended (Toepoel 2012; Singer and Ye 2013; Singer 2018).

Fifth, factors of the fieldwork strategy, such as a larger number of reminders, the number of contact attempts made, or applied refusal conversion affect the response rate positively (Sturgis et al. 2017; Beullens et al. 2018; Klingwort et al. 2018; McGonagle et al. 2022).

Sixth, emphasizing the added value of participation through an appealing display of survey content. This includes adequate explanations regarding confidentiality protection, contact options in case of inquiries, and appropriate communication channels (Dillman 2020).

Seventh, considering the features and attributes of the target population. Individual attributes, for example, education or gender, affect whether a response is given. Accordingly, these characteristics must be considered when surveying a population, especially if they correlate with the target variables of the survey (Dillman 2020).

Of the seven features listed, the data collection mode, the sponsorship/responsible survey conductor, and features of the target population will be considered in the present analysis. In addition, we will use information about the study

year and the coverage area. The discussion will address why other aspects of Dillman (2020) were omitted.

2.1 Crime surveys

Self-report survey methods inform about the “dark figures” or “hidden figures” of crime and are therefore important in criminological and social research. Such surveys complement official crime statistics based on official crime cases reported to the police or cases that lead to an arrest. Combining both data sources allows quantifying crime rates and fear of crime (Maxfield et al. 2000). Methodological aspects in such surveys are discussed by (Kreuter 2002; Noack 2015; Schnell and Noack 2015). Countries such as the USA (National Crime Victimization Survey), the UK (Crime Survey for England & Wales), and the Netherlands (Safety Monitor) conduct periodically repeated victimization surveys of the general population. In Germany, no regular or repeated standardized official crime survey is conducted. From a criminological and national security perspective, this is an unfortunate state of affairs (addressed in the German national security reports (Bundesministerium des Innern and Bundesministerium der Justiz 2001, 2006; Bundesministerium des Innern, für Bau und Heimat and Bundesministerium der Justiz und für Verbraucherschutz 2021)). Individual and independent surveys instead dominate crime survey research. This fact enables us to systematically study the effects of survey design features on the response rate. However, the first studies that use big data for crime statistics are published, but these need to be carefully evaluated (Lohr 2019). Accordingly, sample surveys seem currently without alternatives.

3 Study selection

The initial systematic review was done by Klingwort (2017) and considered the period from 2001–2016. Since then, additional and relevant crime surveys have been conducted. Therefore, the systematic review has been updated for this paper. As the focus is on German victimization surveys, literature research was conducted primarily in German databases and at German institutions. The databases used are: “Deutsche Nationalbibliothek”, “Gemeinsamer Verbundkatalog”, “Sociological Abstracts”, “Sowiport”, “Web of Science”, “WISO”, “Google Scholar”, and “Krim-Dok”. Furthermore, the websites and archives of the following institutions and ministries were queried: “Bundeskriminalamt”, “Landeskriminalämter”, “Bundesministerium für Familie, Senioren, Frauen und Jugend”, “Bundesministerium für Gesundheit”, “Bundesministerium des Innern”, “Statistisches Bundesamt”, “Statistische Lan-

desämter”, “Robert Koch-Institut”, “Leibniz-Institut für Sozialwissenschaften”, “Deutsches Institut für Wirtschaftsforschung”, “Kriminologisches Forschungsinstitut Niedersachsen”, “Eurostat”, “European Union Agency for Fundamental Rights”, “Deutsche Forschungsgemeinschaft”,

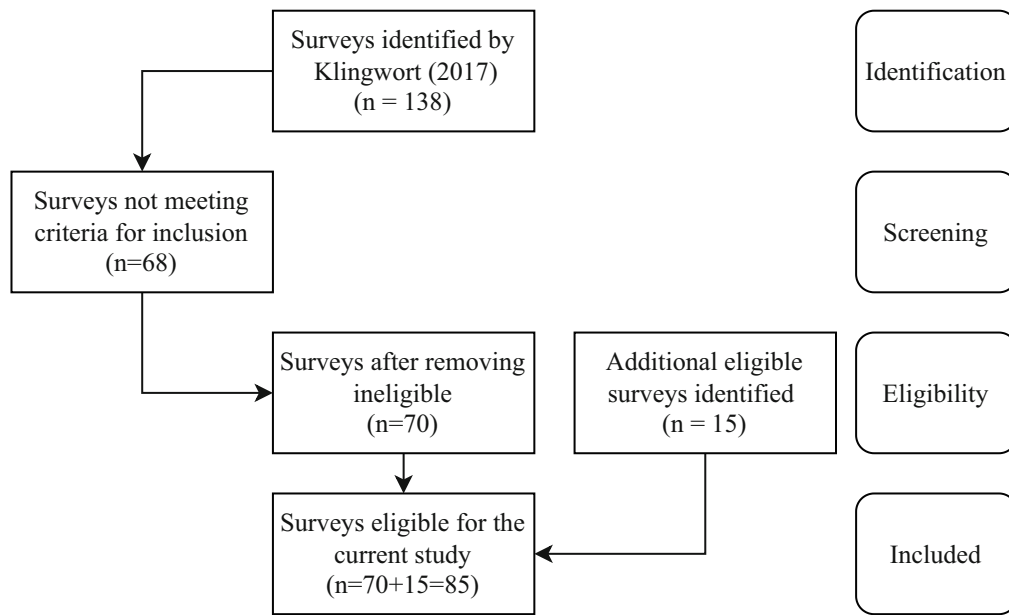
The following German search terms were used: “Opferbefragung (*)”, “Viktimisierungssurvey (*)”, “Viktimisierungsbefragung (*)”, “Kriminalitätsfurcht”, “Befragung zu Kriminalitätsfurcht”, “Befragung zu Wohnungseinbruch”, “Befragung zu Sicherheit”, “Befragung zu Kriminalität”, “Befragung zu Gewalt”. Furthermore the references in the researched sources were searched for further eligible studies. The criteria that the surveys must meet to be included are:

1. Population: Germany
2. Field period: 01.01.2001–31.12.2021
3. Gross sample size: $n \gtrapprox 1000$
4. Questions about forms of physical violence and/or, victimization in the form of residential burglary and/or, questions on fear of crime
5. (Gross) sample size, (net) response size, and response rate

Population “Germany” means that the sample is based on a sampling frame with addresses/contact information within Germany. There are no restrictions regarding German citizenship or the German-speaking population. The chosen field period connects with some overlap to the last overview published by Obergfell-Fuchs (2008) that listed the victimization surveys conducted in Germany between 1973 and 2005. The gross sample size was set $n \gtrapprox 1000$ to prevent a potentially large number of surveys with small sample sizes from being included and to consider only the larger and more relevant surveys. How many more surveys would have been included without this restriction is unknown. However, we explain in this section that two surveys with a smaller gross sample size were included.

Figure 1 shows how the eligible surveys were identified (according to PRISMA guidelines, see <https://prisma-statement.org/>). Klingwort (2017) identified 138 different samples. Note that the number of samples does not correspond to the number of studies. It is possible that more than one independent sample has been drawn within one study, which, for example, considered different target populations. In this case, all independent samples of that study have been considered eligible and the response rates were calculated for each individual sample.

Of the 138 different samples, 68 samples are considered not eligible for the current study. Those are either quota samples not allowing to calculate a response rate, other forms of non-probability samples, a long-term panel, or other general population surveys without a focus on crime. In two cases, drop-off questionnaires were used, which are

**Fig. 1**

PRISMA search strategy to identify eligible surveys

Table 1

Distribution of response rates split by different sets of data

Set	Min.	Q_{25}	Q_{50}	Mean	Q_{75}	Max.
All ($n=85$)	0.127	0.390	0.569	0.580	0.828	0.941
Csd ($n=45$)	0.127	0.270	0.395	0.412	0.511	0.860

not considered in this study because they are not based on an independent sample. Two included studies are below the targeted gross sample size but included because they belong to a study in which an additional independent larger sample was drawn that qualified to be included.

For the current study, additional 15 samples were identified. These studies were identified using the same strategy as for the initial systematic review by Klingwort (2017). However, only the number of eligible surveys is documented. As a result, $n = 85$ samples are eligible and included.

The publications for the identified studies were screened, and reported information about the (gross) sample size, (net) response size, response rate, and the survey design features (study year, target population, coverage area, data

collection mode¹, and institute) were documented. In some cases, only the response size and the response rate were reported. Here, the gross sample size was calculated as $\left\lfloor \frac{\text{response size} \cdot 100}{\text{response rate (in \%)}} \right\rfloor$. The documented variables are described in the following.

3.1 Response rates

We have calculated the response rates using the sample and response size to ensure that the response rates are the same type. By this, we avoid having differently calculated response rates in the target variable. The distribution of the response rates of the samples used is reported in Table 1. The set “All” is based upon all 85 samples with an average response rate of 0.580 and 0.127 as minimum and 0.941 as maximum. Of those 85 samples, 40 are based upon “classroom interviews/School-panels”. Given that this design differs considerably from a cross-sectional design (Csd), we also report the response rate distribution solely for the 45 Csd samples. For the Csd studies, an average response rate of 0.412 is observed, with 0.127 as the minimum and 0.860 as the maximum. In both sets, the mean and median are close, indicating symmetrical distributions

¹ CATI: Computer Assisted Telephone Interview, CAWI: Computer Assisted Web Interview, F2F: Face-to-face interview, PAPI: Paper assisted personal interview

Table 2*Distribution and categorization of survey design features*

Feature	Distribution and categorization
Study year	Min: 2001, Mean: 2010, Max.: 2020
Target population	General population (36); Non-general population (49)
Coverage area	National level (20); Regional level (26); Local level (39)
Data collection mode	CATI (14); CAWI (4); F2F (6); PAPI (21); Classroom interviews/School-panels (40)
Institute	Public institute (10); Ministry (12); National criminological institute (32); Police (15); University (16)

(see also Sect. 4.1). Both sets will be considered for the meta-analysis of proportions, see Sect. 5.

3.2 Survey design features and coding

Table 2 shows the documented survey design features, their distribution and categorization.

4 Methods

4.1 Meta-analysis of proportions

A meta-analysis of proportions considers a proportion as effect size, such as the proportion of people in a study who experienced a particular outcome, e.g., recovering from a disease (Borenstein 2009; Schwarzer and Rücker 2022). Transferred to the current study, we aim at synthesizing response rates from multiple studies to provide a precise estimate of the true effect size, i.e., the response rate. For the current study, it is reasonable to assume that the studies do not stem from a single population, and therefore, a random-effects model will be fitted that assumes a distribution of true effect sizes. Hence, the mean of the distribution of true effect sizes is estimated (Borenstein 2009; Harrer et al. 2022). The random-effects model can be expressed as

$$\hat{\theta}_k = \mu + \zeta_k + \epsilon_k, \quad (1)$$

with $\hat{\theta}_k$ being the observed effect size of study k , μ the true effect sizes mean, and the two error sources ζ_k (true variation in effect sizes) and ϵ_k (sampling error) (Borenstein 2009; Harrer et al. 2022). A weighted mean is computed to

Table 3*Summary proportion of the response rate (true effect size), standard error (SE), and 95% confidence interval (CI) split by different sets of data*

Set	Estimate	SE	CI	
			Lower	Upper
All	0.580	0.026	0.529	0.631
Csd	0.412	0.025	0.364	0.461

obtain a precise estimate of μ , with the weights based on the inverse of a study's variance (sum of the within-study variance and the between-studies variance). The within-study and between-studies variance are necessary to compute a study's variance. Therefore, different estimators exist, with differences in estimating the between-study variance. In the current study, the restricted maximum-likelihood estimator (REML) is used to estimate the between-study variance (Thorlund et al. 2011). The REML is a suitable estimator for continuous outcome data. For details, we refer to Borenstein (2009); Chen and Peace (2013); Harrer et al. (2022); Evangelou and Veroniki (2022). Other available estimators were also tested, such as DerSimonian-Laird, Paule-Mandel, Empirical Bayes, and Sidik-Jonkman. None of these estimators would have yielded substantially different results than those presented in Table 3. The results of this analysis are not shown.

It might be required to apply transformations to the considered proportions to improve their statistical properties regarding following a normal distribution. Usually, the logit or double-arcsine transformation is used (Schwarzer and Rücker 2022). However, when proportions around 0.5 are observed, and the number of studies is large, it can be assumed the proportions follow a binomial distribution. In such a case, no transformations are required because the normal distribution approximates the binomial distribution. Moreover, when the observed proportions are between 0.2 and 0.8, no transformations are required (Lipsey and Wilson 2001). Considering the response rate distributions reported in Table 1, the observed proportions will be used, and no transformations will be applied.

To quantify heterogeneity, we report several metrics. First, Cochran's Q is used to disentangle the sampling error and the between-study heterogeneity. It is used to test if the heterogeneity exceeds the one expected under the null (no heterogeneity). Second, the I^2 statistic, which is the percentage of variability, that is not caused by sampling error. Finally, the heterogeneity variance (τ^2) and the heterogeneity standard deviation (τ). The variance in the true effect sizes is quantified by τ^2 . From that, τ is obtained and is

expressed in the same metric as the effect sizes (Borenstein 2009; Harrer et al. 2022).

4.2 Meta-regression

A meta-regression assumes a mixed-effects model and accounts for the study's deviation from the true effect because of between-study variance and sampling error. In a meta-regression study, design features that may have influenced the results can be used. Hence, a meta-regression predicts the observed effect size $\hat{\theta}_k$ (Harrer et al. 2022). The model is expressed as

$$\hat{\theta}_k = \theta + \beta x_k + \epsilon_k + \zeta_k, \quad (2)$$

with β the regression coefficient and x the predictor (study design feature) of study k . For details, we refer to Borenstein (2009); Harrer et al. (2022). Accordingly, the model in the present study will be based on $k = 85$ data points. Given the limited number of data points, the regression will be restricted to main effects only. For the meta-regression, τ^2 , τ , I^2 , will be reported as well. Moreover, R_*^2 is reported, which considers the residual heterogeneity variance not explained by the meta-regression slope and relates it to the total heterogeneity. Finally, residual heterogeneity and moderators are tested.

All analyses were conducted using R, version 4.1.3 (R Core Team 2022). For the meta-analyses, the R-library metafor was used (Viechtbauer 2010).

5 Results

The estimates of the summary proportion are shown in Table 3. For the set “All”, a summary proportion of 0.580 with an SE of 0.026 and a CI between 0.529 and 0.631 is estimated. For the set “Csd”, a summary proportion of 0.412 with an SE of 0.025 and a CI between 0.364 and 0.461 is estimated. Thus, the random-effects model with the REML estimator yields similar estimates for the summary proportion of the response rate as the means shown in Table 1.

The heterogeneity test for Set “All” shows that $Q = 346,676.8753$, which is larger than expected with $df = 84$. Accordingly, the test for heterogeneity is significant ($p < 0.05$). For the set “Csd”, $Q = 74,352.5739$ and larger than expected with $df = 44$. The test for heterogeneity in this set is also significant ($p < 0.05$). Table 4 shows the additional metrics to quantify heterogeneity. First, the results for set “All” are considered. The between-study heterogeneity variance (τ^2) is estimated at 0.058 (95% CI: 0.044 - 0.081).

Table 4

Metrics to quantify heterogeneity split by different sets of data

	Estimate	CI	
		Lower	Upper
All			
τ^2	0.058	0.044	0.081
τ	0.242	0.210	0.285
I^2 (%)	99.979	99.972	99.985
Csd			
τ^2	0.028	0.019	0.044
τ	0.167	0.138	0.211
I^2 (%)	99.967	99.952	99.980

The true response rates (effect sizes) have an estimated standard deviation at $\tau = 0.242$ (95% CI: 0.210 - 0.285). The I^2 statistic is 99.979% (95% CI: 99.972 - 99.985), indicating substantial heterogeneity. Second, τ^2 for the set “Csd” is estimated at 0.028 (95% CI: 0.019 - 0.044) and $\tau = 0.167$ (95% CI: 0.138 - 0.211). The I^2 statistic is 99.967% (95% CI: 99.952 - 99.980), indicating substantial heterogeneity. Moreover, the confidence intervals for τ^2 of both sets do not contain zero indicating between-study heterogeneity. Summarizing the results from these metrics, there is strong evidence that true response size differences cause variation in the data. The extent to which the considered survey design features are able to explaining the variation in the data will be reported in the next section.

In an additional analysis, we checked whether these reported results and conclusions would change if logit-transformed proportions had been used. No substantial differences were found. The estimates in Table 3 would vary by 1-2 percentage points. The heterogeneity measures reported in Table 4 would yield the same conclusions.

5.1 Meta-regression

Table 5 shows the results of the meta-regression. A permutation test for the p-values is used to validate the robustness of the model. By this, it is possible to assess better whether the coefficients capture a true pattern or if the model picks up statistical noise (see also Harrer et al. (2022)).

The intercept is estimated at 22.835 (95% CI: 9.179 - 36.491) with an SE of 6.967. The feature “Year” has a negative coefficient (-0.011), and the CI indicates only negative effects of this feature. Accordingly, on average, each progressing year causes a decrease in the response rate of -0.011. The p-value of 0.002 indicates statistical signifi-

Table 5

Results of the Meta-regression model. The estimated coefficients, standard error (SE), and confidence interval (CI) are shown. The p-values are based on a permutation test with 1000 iterations

	β	SE	p	CI	
				Lower	Upper
Intercept	22.835	6.967	0.002	9.179	36.491
Year	-0.011	0.003	0.002	-0.018	-0.004
Population (reference: general population)					
Non-general pop.	-0.045	0.052	0.404	-0.146	0.056
Coverage area (reference: local)					
National	0.103	0.051	0.055	0.004	0.203
Regional	0.022	0.040	0.608	-0.057	0.101
Data collection (reference: CATI)					
CAWI	0.164	0.083	0.044	0.001	0.327
Classroom inter-views/School-panels	0.444	0.076	0.001	0.296	0.593
F2F	0.137	0.081	0.088	-0.022	0.296
PAPI	0.201	0.051	0.001	0.101	0.301
Institute (reference: Ministry)					
National crim. institute	0.049	0.057	0.413	-0.063	0.161
Police	-0.038	0.055	0.495	-0.145	0.069
Public institute	-0.075	0.063	0.222	-0.199	0.048
University	0.180	0.069	0.013	0.045	0.315

$k = 85$; $\tau^2 = 0.017$ ($SE = 0.003$); $I^2 = 99.900$; $R^2_* = 70.310$; $QE = 67,271.114$, $p < 0.05$; $QM = 210.453$, $p < 0.05$

cance, i.e., the null hypothesis can be rejected, and the coefficient is significantly different from 0. Considering the target population, studies surveying non-general populations score on average a -0.045 lower response rate compared with those surveying the general population. The p-value of 0.404 indicates no statistical significance, i.e., the null hypothesis cannot be rejected, and the coefficient is not significantly different from 0. When the survey is conducted nationwide or on a regional level, in both cases, average larger response rates are achieved than with locally conducted surveys. The p-values of 0.055 and 0.608 indicate no statistical significance. All data collection modes score, on average, higher than CATI. For example, using PAPI instead of CATI will, on average, result in an 0.201 higher response rate. When conducting a classroom interview/school-panel, a 0.444 larger response rate can be achieved on average. The p-values of all coefficients indicate statistical significance, except the coefficient for “F2F” (0.088). The coefficients for the feature “Institute” shows that in surveys in which the Police and Public Institutes occupy a prominent function, on average lower response rates are achieved in

contrast to surveys in which a Ministry holds a central role. National criminological institutes and Universities achieve, on average, a higher score than Ministries. Here, only the p-value for the coefficient for “University” (0.013) indicates statistical significance.

The variance that the predictors do not explain in this model is $\tau^2 = 0.017$ ($SE = 0.003$). According to I^2 , after the inclusion of the predictors, 99.900% of the variability is due to the remaining between-study heterogeneity. The R^2_* indicates that the predictors explain 70.310% of the difference in response rates. The test for residual heterogeneity is significant ($p < 0.05$), with $QE = 67271.114$, meaning that the heterogeneity not explained by the predictors is significant. The test of moderators is significant ($p < 0.05$), with $QM = 210.453$, meaning that some of the included predictors influence the study’s response rate.

5.2 Model predictions

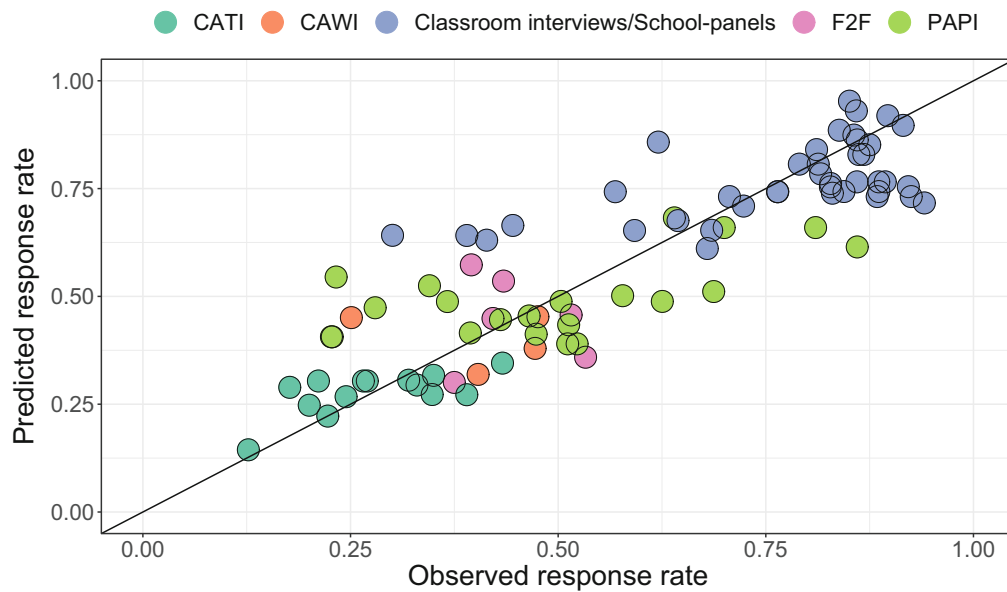
Figure 2 shows the observed and predicted response rate using the observed data and the developed model. The data points are colored by data collection mode and scattered around the regression line, indicating a strong positive correlation ($r = 0.863$).

Table 6 shows the distribution of the absolute prediction errors (observed - predicted). On average, the error is 0. The median is close to zero, indicating a symmetrical distribution. The minimum and maximum indicate large prediction errors of -0.341 (overestimation) and 0.245 (underestimation). The MSE = 0.015 and the RMSE = 0.121.

The developed model can be used for decision-making when (re-) designing a survey. In Table 7, we report the five survey designs with the highest response rates and the five with which the lowest response rates are predicted. Here we do not consider “Classroom interviews/School-panels” as data collection mode and “University” as institute. This decision is due to the use of main effects only. A detailed explanation is given in the limitations of this study, see Sect. 6). The year of conducting the survey is constant in the different designs and is set to 2024. Increasing the year would result in lower response rates, given the negative coefficient.

The results provide evidence that higher response rates are achieved in nationwide surveys where the general population is considered. The use of PAPI for data collection appears to be superior, and CATI surveys will achieve low response rates. With a criminological institute as the responsible institute, the highest response rates can be expected, while surveys conducted by the police or public institutes will achieve low response rates when using CATI.

For the predicted response rates reported, standard errors between 0.068 – 0.106 are found. The range of the reported

**Fig. 2**

Observed and predicted response rates. The color indicates the data collection mode. The black diagonal shows the regression line

Table 6

Distribution of differences between observed and predicted response rates

Min.	Q_{25}	Q_{50}	Mean	Q_{75}	Max.
-0.341	-0.047	0.015	0.000	0.078	0.245

confidence intervals can be considered large. This is likely due to the small number of data points in the model. Moreover, some negative values are predicted. This is due to modeling the original proportions. In these scenarios, the response rate falls between 0 and the upper limit of the confidence interval.

5.3 Model selection

Meta-regressions are usually based on little observations limiting the complexity of the statistical model. Choosing a too-complex model will result in overfitting and imprecise estimates. We aimed to find a probably more sparse model than the model reported in Sect. 5.1 that might still explain a comparable amount of heterogeneity. Therefore, we fitted all potential model combinations (considering main effects only) and compared the obtained R^2_* . When leaving out the population feature, an $R^2_* = 70.41\%$ was obtained. Hence, this might be considered a feature to be dropped when the number of included studies is limited. The worst

model ($R^2_* = 2.9\%$) was obtained using only the study year. The feature itself, strictly speaking, is not an actual survey design feature.

6 Discussion

This section highlights the study findings, addresses its limitations, and gives recommendations for future research.

First, the results of the meta-analysis of proportions showed an estimated summary response rate of 58% and 41.2% when considering cross-sectional designs only. Second, the results of the meta-regression showed that time has a negative effect on the response rate. Surveying the non-general population has a negative effect as well. Surveys on the national or regional level score higher response rates than locally conducted surveys. All data collection modes score higher than CATI. The findings for the effect of the responsible institute are mixed. Some institutes score higher than ministries (national criminological institute, university), and others lower (police, public institutes). The model predictions based on the observed data are of reasonably good quality. However, some of the predictors were not significant. When the model is applied to new data, results suggest using nationwide general population PAPI surveys. It is not recommended to survey non-general populations with CATI on a local or regional level, with the police or a public institute being responsible.

Table 7

Selection of survey design feature sets and their predicted response rate (RR), standard error (SE), and confidence interval (CI)

Population	Coverage	Data collection	Institute	RR	SE	CI	
						Lower	Upper
General	National	PAPI	Crim. institute	0.457	0.068	0.325	0.590
General	National	CAWI	Crim. institute	0.420	0.097	0.230	0.610
Non-general	National	PAPI	Crim. institute	0.413	0.089	0.240	0.586
General	National	PAPI	Ministry	0.409	0.081	0.250	0.567
General	National	F2F	Crim. institute	0.393	0.095	0.207	0.580
Non-general	Regional	CATI	Police	0.043	0.079	-0.112	0.197
General	Local	CATI	Public institute	0.028	0.091	-0.149	0.206
Non-general	Local	CATI	Police	0.021	0.086	-0.148	0.189
Non-general	Regional	CATI	Public institute	0.006	0.103	-0.196	0.207
Non-general	Local	CATI	Public institute	-0.016	0.106	-0.223	0.190

The developed model is easy to adopt, given the limited number of features and their resolution. However, the sparsity of the model also implies some limitations.

First, given the number of data points, the precision of the obtained point estimates is limited. This effect is evident for the estimate precision in Table 3, where the CI ranges are 10%. This problem becomes even more evident in Table 5. The CIs of the coefficients are large and do sometimes not allow clear conclusions in one direction. If the obtained precision is acceptable might be related to the research question or the quality standards. This problem can only be solved by including more studies (data points). A study not limited to crime surveys should be able to find more observations more easily.

Second, although the model explains a significant part of the heterogeneity (about 70%) in the response rates, about a third is not explained, which still contains a significant part of heterogeneity. This is because several potentially important features are not included in the model. For example, the fieldwork duration and strategy, recruitment letter and the number of reminders, survey length, incentives, presence of an interviewer, or interviewer workload. This information could not always be retrieved from the publications or was reported too vague. However, as mentioned, the number of data points is limited, and even if all these features had been available, not all could likely have been included in the model, and model selection would have been required.

Third, we considered only the main effects while using interactions would have been desirable. For example, the institute's case has a problem of using only main effects. The category "University" has a large positive coefficient. However, all observations with "University" also fall into the category "Classroom interviews/School-panels", which

explains the large coefficient. This result is thus somewhat misleading and would be explained by interactions.

Fourth, the generalization of this model to other data and non-crime surveys. It is questionable to what extent the cultural context (e.g., the use of PAPI in Germany) can be transferred to other surveys. However, replication in other countries or with non-crime surveys would provide information about the model's generalizability.

In contrast to several published studies, recent research suggests increasing responderates Holtom et al. (2022). Considering 1014 surveys from 2010 to 2020 a steadily increasing average response rate is reported: 48% (2005), 53% (2010), 56% (2015), and 68% (2020). Considering these findings and the findings of our work, we see these results as an encouragement for survey researchers to work towards well-suited survey designs that will result in higher response rates.

7 Conclusion

We studied the effect of survey design features (nonsampling errors) on the response rates in crime surveys. Heterogeneity in response rates has been observed that cannot be explained by a random process. The survey design features (study year, target population, coverage area, data collection mode, and responsible institute) explained a large part of the observed heterogeneity. Results highlight the need for an appropriate survey design and professional administration.

Acknowledgements We thank two anonymous reviewers for their time, effort, and comments that improved the manuscript.

The views expressed in this report are those of the authors and do not necessarily correspond to the policies of Statistics Netherlands. The code and data are available from the corresponding author upon reasonable request.

References

- Allison, P.D. (2002). *Missing data*. SAGE.
- Alwin, D.F. (2013). Reflections on thirty years of methodology and the next thirty. *Bulletin of Sociological Methodology/Bulletin de Methodologie Sociologique*, 120(1), 28–37.
- Axinn, W.G., Link, C.F., & Groves, R.M. (2011). Responsive survey design, demographic data collection, and models of demographic behavior. *Demography*, 48(3), 1127–1149.
- Beullens, K., Loosveldt, G., Vandenplas, C., & Stoop, I. (2018). *Response rates in the european social survey: Increasing, decreasing, or a matter of fieldwork efforts?* Survey Methods: Insights from the Field. <https://doi.org/10.13094/SMIF-2018-00003>.
- Blumenberg, C., & Barros, A.J.D. (2018). Response rate differences between web and alternative data collection methods for public health research: A systematic review of the literature. *International Journal of Public Health*, 63(6), 765–773.
- Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to meta-analysis*. Wiley.
- Brick, J.M., & Williams, D. (2012). Explaining rising non-response rates in cross-sectional surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1), 36–59. <https://doi.org/10.1177/0002716212456834>.
- Bundesministerium des Innern, & Bundesministerium der Justiz (2001). First periodic safety report [in german: Erster periodischer Sicherheitsbericht].
- Bundesministerium des Innern, & Bundesministerium der Justiz (2006). Second periodic safety report [in german: Zweiter periodischer Sicherheitsbericht].
- Bundesministerium des Innern, für Bau und Heimat, & Bundesministerium der Justiz und für Verbraucherschutz (2021). *Third periodic safety report [in german: Dritter periodischer Sicherheitsbericht]*
- Chen, D.-G., & Peace, K.E. (2013). *Applied meta-analysis with R*. CRC.
- Chun, A.Y., Heeringa, S.G., & Schouten, B. (2018). Responsive and adaptive design for survey optimization. *Journal of Official Statistics*, 34(3), 581–597. <https://doi.org/10.2478/jos-2018-0028>.
- Couper, M.P. (2013). Is the sky falling? new technology, changing media, and the future of surveys. *Survey Research Methods*, 7(3), 145–156.
- Couper, M.P. (2017). New developments in survey data collection. *Annual Review of Sociology*, 43(1), 121–145.
- Czajka, J.L., & Beyler, A. (2016). *Declining response rates in federal surveys: Trends and implications*. (1)
- Daikeler, J., Bosnjak, M., & Manfreda, L.K. (2019). Web versus other survey modes: An updated and extended meta-analysis comparing response rates. *Journal of Survey Statistics and Methodology*, 8(3), 513–539. <https://doi.org/10.1093/jssam/smz008>.
- Daikeler, J., Bosnjak, M., & Manfreda, L.K. (2020). Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates. *Journal of Survey Statistics and Methodology*, 8(3), 513–539. <https://doi.org/10.1093/jssam/smz008>.
- Dillman, D.A. (2020). Towards survey response rate theories that no longer pass each other like strangers in the night. In P.S. Brenner (Ed.), *Understanding survey methodology: Sociological theory and applications* (pp. 15–44). Springer.
- Dillman, D.A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B.L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research*, 38(1), 1–18. <https://doi.org/10.1016/j.ssres>.
- Dillmann, D.A., Smyth, J.D., & Christian, L.M. (2014). *Internet, phone, mail and mixed-mode surveys: The tailored design method* (4th edn.). Wiley.
- Dutwin, D., & Buskirk, T.D. (2021). Telephone Sample Surveys: Dearly Beloved or Nearly Departed? Trends in Survey Errors in the Era of Declining Response Rates. *Journal of Survey Statistics and Methodology*, 9(3), 353–380. <https://doi.org/10.1093/jssam/smz044>.
- Edwards, M.L., Dillman, D.A., & Smyth, J.D. (2014). An experimental test of the effects of survey sponsorship on internet and mail survey response. *Public Opinion Quarterly*, 78(3), 734–750. <https://doi.org/10.1093/poq/nfu027>.
- Evangelou, E., & Veroniki, A.A. (Eds.). (2022). *Metaresearch*. Springer.
- Faria, A.J., & Dickinson, J.R. (1996). The effect of reassured anonymity and sponsor on mail survey response rate and speed with a business population. *Journal of Business & Industrial Marketing*, 11(1), 66–76. <https://doi.org/10.1108/08858629610112300>.
- Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73(2), 349–360.
- Galesic, M., Bruine de Bruin, W., Dalege, J., Feld, S.L., Kreuter, F., Olsson, H., Prelec, D., Stein, D.L., & van der Does, T. (2021). Human social sensing

- is an untapped resource for computational social science. *Nature*, 595(7866), 214–222. <https://doi.org/10.1038/s41586-021-03649-2>.
- Greenlaw, C., & Brown-Welty, S. (2009). A Comparison of Web-Based and Paper-Based Survey Methods: Testing Assumptions of Survey Mode and Response Cost. *Eval. Rev.*, 33(5), 464–480. <https://doi.org/10.1177/0193841X09340214>.
- Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society. Series A*, 169(3), 439–457.
- Habermann, H., Kennedy, C., & Lahiri, P. (2017). A conversation with robert groves. *Statistical Science*, 32(1), 128–137.
- Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2022). *Doing meta-analysis with R: A hands-on guide*. CRC Press.
- Holtom, B., Baruch, Y., Aguinis, H., & Ballinger, A. G. (2022). Survey response rates: Trends and a validity assessment framework. *Human Relations*, 75(8), 1560–1584. <https://doi.org/10.1177/00187267211070769>.
- Isani, M., & Schlipphak, B. (2022). Who is asking? the effect of survey sponsor misperception on political trust: Evidence from the afrobarometer. *Quality & Quantity*. <https://doi.org/10.1007/s11135-022-01517-3>.
- Klingwort, J. (2017). Nonresponse in current german victimization surveys [in german: Nonresponse in aktuellen deutschen Viktimisierungssurveys. *Duisburger Beiträge zur soziologischen Forschung*. <https://doi.org/10.6104/DBsF-2017-0>.
- Klingwort, J., Buelens, B., & Schnell, R. (2018). Early versus late respondents in web surveys: Evidence from a national health survey. *Statistical Journal of the IAOS*, 34(3), 461–471.
- Kohut, A., Keeter, S., Doherty, C., Dimock, M., & Christian, L. (2012). *Assessing the representativeness of public opinion surveys*. The Pew Research Center.
- Kreuter, F. (2002). *Kriminalitätsfurcht: Messung und methodische Probleme*. VS.
- Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Sacrificing in surveys: Initial evidence. *New Directions for Evaluation*, 1996(70), 29–44. <https://doi.org/10.1002/ev.1033>.
- Laflamme, F., & Wagner, J. (2016). Responsive and adaptive designs. In C. Wolf, D. Joye, T. W. Smith & Y. C. Fu (Eds.), *The SAGE handbook of survey methodology* (pp. 397–408). SAGE.
- de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, 21(2), 233–255.
- Link, M. W., Murphy, J., Schober, M. F., Buskirk, T. D., Childs, H. J., & Tesfaye, L. C. (2014). Mobile technologies for conducting, augmenting and potentially replacing surveys. executive summary of the aapor task force on emerging technologies in public opinion research. *Public Opinion Quarterly*, 78(4), 779–787. <https://doi.org/10.1093/poq/nfu054>.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical metaanalysis*. SAGE.
- Lohr, S. L. (2019). *Measuring crime: Behind the statistics*. CRC Press.
- Lugtig, P., Struminskaya, B., Timmers, A., Henneveltdt, C., Peytcheva, E., & Groves, R. (2022). *The relation between nonresponse rates and nonresponse bias. an update and extension of groves and peytcheva (2008)*. Chicago: AAPOR conference - May 10-14, 2022.
- Luiten, A., & Schouten, B. (2013). Tailored fieldwork design to increase representative household survey response: An experiment in the survey of consumer satisfaction. *Journal of the Royal Statistical Society. Series A*, 176(1), 169–189.
- Luiten, A., Hox, J., & de Leeuw, E. (2020). Survey Non-response Trends and Fieldwork Effort in the 21st Century: Results of an International Study across Countries and Surveys. *Journal of Official Statistics*, 36(3), 469–487. <https://doi.org/10.2478/jos-2020-0025>.
- Maxfield, M. G., Weiler, B. L., & Widom, C. S. (2000). Comparing self-reports and official records of arrests. *Journal of Quantitative Criminology*, 16(1), 87–110.
- Mcgonagle, K. A., Sastry, N., & Freedman, V. A. (2022). The effects of a targeted “early bird” incentive strategy on response rates, fieldwork effort, and costs in a national panel study. *Journal of Survey Statistics and Methodology*, smab042 <https://doi.org/10.1093/jssam/smab042>.
- Meyer, B. D., Mok, W. K., & Sullivan, J. X. (2015). Household surveys in crisis. *Journal of Economic Perspectives*, 29(4), 199–226.
- Millar, M. M., & Dillmann, D. A. (2011). Improving response rate to web and mixed-mode surveys. *Public Opinion Quarterly*, 75(2), 249–269. <https://doi.org/10.1093/poq/nfr003>.
- Noack, M. (2015). *Methodische Probleme bei der Messung von Kriminalitätsfrucht und Viktimisierungserfahrungen*. Springer VS.
- Obergfell-Fuchs, J. (2008). Crime victims and insecurity surveys in germany. In R. Zauberman (Ed.), *Victimisation and insecurity in europe: A review of surveys and their use* (pp. 105–125). University Press.

- Olson, K., Smyth, J.D., Horwitz, R., Keeter, S., Lesser, V., Marken, S., Mathiowetz, N.A., McCarthy, J.S., O'Brien, E., Opsomer, J.D., Steiger, D., Sterrett, D., Su, J., Suzer-Gurtekin, Z.T., Turakhia, C., & Wagner, J. (2020). Transitions from telephone surveys to self-administered and mixed-mode surveys: Aapor task force report. *Journal of Survey Statistics and Methodology*, 9(3), 381–411. <https://doi.org/10.1093/jssam/smz062>.
- Presser, S., & McCulloch, S. (2011). The growth of survey research in the united states: Government-sponsored surveys, 1984–2004. *Social Science Research*, 40(4), 1019–1024.
- R Core Team (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rao, J., & Fuller, W.A. (2017). Sample survey theory and methods: Past, present, and future directions. *Survey Methodology*, 43(2), 145–160.
- Schnell, R. (2019). *Survey-interviews: Methoden standardisierter befragungen* (2nd edn.). Springer VS.
- Schnell, R., & Noack, M. (2015). Stichproben, nonresponse und gewichtung für viktimisierungsstudien. In N. Guzy, C. Birkel & R. Mischkowitz (Eds.), *Methodik und Methodologie*. Viktimisierungsbefragungen in deutschland, (Vol. 2, pp. 8–75). Bundeskriminalamt.
- Schouten, B., Calinescu, M., & Luiten, A. (2013). Optimizing quality of response through adaptive survey designs. *Survey Methodology*, 39(1), 29–58.
- Schwarzer, G., & Rücker, G. (2022). Meta-analysis of proportions. In E. Evangelou & A.A. Veroniki (Eds.), *Meta-research: Methods and protocols* (p. 159172). Springer. https://doi.org/10.1007/978-1-0716-1566-9_10.
- Singer, E. (2016). Reflections on surveys' past and future. *Journal of Survey Statistics and Methodology*, 4(4), 463–475.
- Singer, E. (2018). The use and effects of incentives in surveys. In D.L. Vannette & J.A. Krosnick (Eds.), *The palgrave handbook of survey research* (pp. 63–70). Palgrave Macmillan.
- Singer, E., & Ye, C. (2013). The use and effects of incentives in surveys. *Annals of the American Academy of Political and Social Science*, 645(1), 112–141. <https://doi.org/10.1177/0002716212458082>.
- Sturgis, P., Williams, J., Brunton-Smith, I., & Moore, J. (2017). Fieldwork effort, response rate, and the distribution of survey outcomes: a multilevel meta-analysis. *Public Opinion Quarterly*, 81(2), 523–542. <https://doi.org/10.1093/poq/nfw055>.
- Thorlund, K., Wetterslev, J., Awad, T., Thabane, L., & Gluud, C. (2011). Comparison of statistical inferences from the dersimonian-laird and alternative random-effects model meta-analyses – an empirical assessment of 920 cochrane primary outcome meta-analyses. *Research Synthesis Methods*, 2(4), 238–253. <https://doi.org/10.1002/jrsm.53>.
- Toepoel, V. (2012). Effects of incentives in surveys. In *Handbook of survey methodology for the social sciences* (pp. 209–223). Springer.
- Tourangeau, R., & Smith, T.W. (1996). Asking sensitive questions: the impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, 60(2), 275–304. <https://doi.org/10.1086/297751>.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>.
- Williams, D., & Brick, J.M. (2018). Trends in U.S. Face-To-Face Household Survey Nonresponse and Level of Effort. *Journal of Survey Statistics and Methodology*, 6(2), 186–211. <https://doi.org/10.1093/jssam/smx019>.
- Yan, T., & Williams, D. (2022). Response burden – review and conceptual framework. *Journal of Official Statistics*, 38(4), 939–961. <https://doi.org/10.2478/jos-2022-0041>.
- Yan, T., Fricker, S., & Tsai, S. (2020). Response burden: What is it and what predicts it? In *Advances in questionnaire design, development, evaluation and testing* (pp. 193–212). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119263685.ch8>.