Survey Research Methods (2023)
Vol. 17, No. 4, pp. 465-491
https://doi.org/10.18148/srm/2023.v17i4.8126
European Survey Research Association

©2023 Author(s)
CC BY 4.0

Late Responding in Web and Mail Surveys: A Systematic Review and Meta-Analysis

Ellen Laupper[1,2] · Esther Kaufmann[1] · Reips Ulf-Dietrich[1]
[1]iScience Group, Department of Psychology, University of Konstanz
[2]Swiss Federal University for Vocational Education and Training SFUVET

There is a fundamental concern that respondents who complete a survey with a certain delay or only after one or more additional contact attempts are less motivated to provide high quality survey data. Given the rise of web surveys to being the currently most widely used mode of the survey method, this concern has increased, and surveyors wonder which mode to choose best. With a systematic review and a meta-analytic approach, we clarify types of and issues in "late responding", and we address the questions of whether and to what extent late responding is different for web surveys compared to mail surveys. The systematic review reveals that only a third of the 74 studies included report on data quality for any type of late responding. Moreover, a wide range of definitions for late responding was identified, with essentially three types. With a meta-analytical approach, a mean share of 27% (CI: 23%–31%) of late responding across both modes was quantified, and no mode difference was found. A moderator analysis with 16 sample and survey characteristics did not identify a robust moderator across modes. In addition, our article provides a detailed overview of different survey practices used in web and mail surveys.

Keywords: late respondents; systematic review; meta-analysis; survey research; web survey; mode effect; late responding

1 Introduction

As the use of web surveys over the last decades has become increasingly ubiquitous in research (e.g., social and market research), research on respondents' completion behavior and data quality in web surveys has provided many detailed insights (e.g., Callegaro et al., 2015; Cornesse & Bosnjak, 2018; Crews & Curtis, 2011; Dillman, 2021; Dodou & De Winter, 2014; Morrison, 2011; Venette et al., 2010). Systematic reviews and aggregations of study results through meta-analyses focusing on methodological comparisons of the web survey with other survey modes reveal that web surveys appear to, for example, be less representative (Cornesse & Bosnjak, 2018) and have, on average, about 11% lower response rates (Daikeler et al., 2019; Shih & Fan, 2007). Whereas, the average item nonresponse rates for web surveys are comparable to those of other survey modes (Čehovin et al., 2022), and social desirability scores are the same as in other self-administered surveys (Dodou & De Winter, 2014; Gnambs & Kaspar, 2017, but see Kaufmann & Reips, 2008). Considering these substantiated errors in coverage and unit nonresponse of the web survey mode another important and related issue, on which we do not know whether it is more or less pronounced for the web survey mode, is "late responding".

Corresponding author: Ellen Laupper, iScience Group, Department of Psychology, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany (Email: ellen.laupper@sfuvet.swiss)

2 Current Research On Late Responding

Late responding refers to study participants who complete the survey with a certain delay or only after one or more additional contact attempts (see Olson, 2013). There is a fundamental concern that these respondents pose a threat to the accuracy of survey estimates, as late respondents appear to provide survey data of poorer quality (Brehaut et al., 2006; Helasoja et al., 2002; Kaminska et al., 2010; Korkeila et al., 2001; Kreuter et al., 2014; Olson, 2013; Peytchev & Hill, 2010; Yan et al., 2004). On the other hand, there is a substantial body of evidence indicating that including data from late respondents increases the accuracy of survey estimates by improving the representativity of the sample. Late and early respondents differ in sociodemographic characteristics, such as late respondents being younger, having a lower educational level, being from a lower socioeconomic strata, having a minority status, reporting a poorer health status,

and being male (e.g., Chen et al., 2003; Klingwort et al., 2018; Manjer et al., 2002; Sigman et al., 2014; Vink & Boomsma, 2008). They also differ in relevant study outcomes, such as measures of satisfaction, attitudes, or behaviors (e.g., Chen et al., 2003; Estelami, 2015; Goos & Salomons, 2017; Kypri et al., 2011; Manjer et al., 2002; Vink & Boomsma, 2008; Voigt et al., 2003). The extent to which efforts to motivate late respondents to complete the survey affect data quality remains an unresolved question.

2.1 Definition

Up to now, there is only one systematic review on late responding. In her review Olson (2013) defines late responding as answering a survey late or after several encouragements. This generic definition touches on two basic questions concerning the understanding and definition of the phenomenon and its relationship to data quality and nonresponse bias. On a fundamental level, it leaves open whether late responding is a continuous or a discontinuous phenomenon, linked only to the passing of time or linked to one or several specific events in the survey administration process, such as one or several contact attempts. Specifically, no further information is provided about what sort of encouragement or which number of encouragements divides respondents into early and late. It seems that late responding is a rather vague phenomenon, and as survey designs differ substantially in their strategies to increase response rates, it is not surprising that the operationalization of the phenomenon, the share of late respondents, and its impact on data quality and nonresponse bias found in the literature vary to a similar extent (Bach et al., 2020). Moreover, only few studies explicitly elaborate on their chosen definition of the phenomenon. For example, Oremus and Wolfson (2004) argue that their late respondents would have been nonrespondents without the follow-ups. Vink and Boomsma (2008) base their operationalization on the observation that the number of subjects returning the survey reaches an asymptote after Day 30. Whereas, Friedman et al. (2003; p. 993) reason that "any survey arriving after the cut-off for its respective fielding period was deemed a late respondent, for it was too late to be included in the quarterly dataset". Meanwhile, Sigman et al. (2014) tested several thresholds (e.g., after the first half of the fielding period vs. after the first month), and after observing that for most of these operationalizations, almost 100% were "early respondents" they agreed on a two-week threshold. Given this ambiguous starting point, it is important first to have an overview of the range of operationalizations used in the relevant research literature. In particular, it is important to know whether there are certain preferences and, if so, whether they differ between the two modes.

2.2 Reasons

To reduce late responding, it is relevant to know why participants respond late. Although Olson (2013) found only a few studies that measured the potential causes of late responding and linked it to data quality, she found some empirical support for the *general lack of motivation, reactance*, and *lack of interest in the topic hypotheses* resulting in higher item nonresponse. On the other hand, she found no such support for the *increase in perceived importance*, predicting lower item nonresponse with an increase in contact attempts. Regarding the few existing indications as to why participants respond late, one could also argue that most of the identified reasons for nonresponses can be considered the same for late responding. These include psychological reactance (Reips, 2013), not being interested in the study or doubting the usefulness of study outcomes, survey fatigue, misplacing the questionnaire, lack of time (e.g., too busy, vacation), forgetfulness, and confidentiality concerns (e.g., Dahlhamer et al., 2008; Emberton & Black, 1995; Keusch, 2015; Olson et al., 2019; Olson et al., 2010). Thus, late responding has a motivational component.

2.3 Moderators

As far as we know, there is little research on which survey (design) features lead to late responding. Again, assuming that there is some relationship between delayed response and nonresponse, it seems plausible to expect that variables that cause nonresponse also play a role in delayed response. Dillman (2021) summarizes seven factors that are known to likely influence response rates: (1) multiple modes, (2) sponsorship, (3) survey length, (4) incentives, (5) number and timing of survey requests, (6) appealing and less burdensome surveys, and (7) personal characteristics of the target population. In their meta-analysis, Groves and Peytcheva (2008) identified the following survey design variables as having an impact on the relationship between nonresponse rates and nonresponse bias: type of sponsorship (government vs. other) and whether or not respondents had prior interaction with the sponsorship as well as self- vs. interview-administered survey modes. Moreover, they found that surveys of the general population tend to yield larger nonresponse rates than surveys of more specific populations. By including potentially relevant sample and survey design factors in our research, we may identify which sample design characteristics are responsible for any differences in late responding between web and mail surveys.

2.4 Web survey mode and late responding

We focus specifically on the web survey mode in our study. This is because it is well established that coverage and non-response errors are different for web surveys than for other survey modes. Moreover, the only systematic review on this topic by Olson (2013) only considers interview and mail surveys. In addition, different modes offer different survey design options.

Web surveys offer participants much flexibility and ease to respond to a survey request with just a few mouse clicks. However, such a request sent by email is perceived as less committal than, for example, a personal request via telephone, face-to-face or mail (Dillman et al., 2010; Keusch, 2015). In addition to having different implications for respondents in terms of their perception of cognitive burden, privacy, legitimacy, ease of answering, and others (e.g., Dillman et al., 2014; Tourangeau et al., 2000), different survey modes offer various design options and costs and employ disparate contact strategy practices. Consequently, it is likely that there are also differences between these modes in terms of how they can persuade (reluctant) respondents to participate. Therefore, we expect mode to be one of the most important moderators for late responding.

For comparability, we focus in our comparison on the two self-administered modes web and mail survey. This is because interview-administered surveying adds with the interviewers an additional source of variance.

3 Research Questions

As outlined, an in-depth understanding of the late-responding phenomenon and its contributing factors is relevant for survey research to generate more knowledge on how to increase the accuracy of survey estimates.

However, late responding is ambiguous in how it is defined and operationalized. To fill these research gaps, we conduct a systematic review and meta-analysis of late responding. We are especially interested in a comparison of the web survey mode to self-administered surveys conducted on paper-and-pencil, as mode should be considered a relevant source of variability and no comprehensive knowledge on the late-responding phenomenon exists to date for this currently most widely used mode. With this systematic review, we first assess the range of definitions used in the relevant studies and record the differences in study characteristics possibly relevant to moderating late responding. In the second step, we assess the proportions of late responding and thirdly identify possible moderator variables pertinent to the phenomenon with a meta-analytic approach. Hence, with the focus on the two self-administered modes—web and mail surveys, this article addresses the following three research questions:

(1) How is late responding operationalized? Are there differences in definitions used between web surveys and mail surveys?
(2) What is the proportion of late respondents for web surveys compared to mail surveys?
(3) Are survey characteristics (e.g., target population/sample or survey design) moderating the proportion of late respondents? Are survey characteristics moderating the proportion of late respondents differently for web and mail surveys?

4 Method

To answer our research questions, we use a best-evidence synthesis approach (Slavin, 1986) that combines a systematic review and a meta-analytic approach.

4.1 Systematic Review

Our systematic review is composed of three steps (Siddaway et al., 2019). First, we conduct a systematic literature research following the PRISMA (preferred reporting items for systematic reviews and meta-analyses) guidelines (Liberati et al., 2009).[1] Second, we review and select the records identified by the literature review according to our predefined eligibility criteria. Third, we obtain the manuscripts and code the information on potentially relevant study design, sample, and target population characteristics as potential moderator variables and outcome variables (i.e., late respondent definition, number, and percentage of late respondents).

4.1.1 Eligibility Criteria and Search Strategy

For the systematic review, we focus on web and mail survey studies that reported the proportion of late respondents. We do not impose restrictions on the type of literature, field of research, study topic, or type of study outcome (e.g., satisfaction, attitude, and behavior), but we establish the following five eligibility criteria:

(1) Because web pioneers did not send out their first online surveys until the mid-1990s (Reips, 2021), we considered studies published between 2000 and March 2021.

[1] PRISMA is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses (see https://prisma-statement.org/).

Table 1

Used Search Terms and Search Engines

Search terms	"survey" or "questionnaire" combined with "late respond*"or "reluctant respond*" (combined with "data quality" or "item nonresponse" or "item non-response" or "missing*" for the search in the VULibrary)
Search engines	VULibrary bibliographic harvester (Digital Library of the Vrije Universiteit Amsterdam), which includes searching 153 bibliographic data bases like e.g., ProQuest, PsycARTICLES, PubMed, ScienceDirect, Emerald (e.g., Business, Social Sciences), Wiley Online Library, Sage Journals, SpringerLink, EBSCO, and JSTOR Business. Additionally, PsycInfo, ERIC and Web of Science were searched

(2) To avoid any English language bias (see Dickersin, 2005), we also considered non-English, specially German language manuscripts.
(3) Because more direct methods of persuasion can be used in interview-administered survey methods, we excluded studies using survey methods such as telephone or face-to-face interviews and searched for studies reporting on self-administered surveys.
(4) We also excluded Internet and paper-and-pencil surveys that were conducted under controlled conditions in a laboratory or an on-site setting, and web surveys with no well-defined sampling frame, such as surveys advertised through online banners, offline media, or pop-up intercept surveys (e.g., Keusch, 2015).
(5) Except for surveys targeting children or teens under the age of 18, there are no further restrictions for the study population.

As a first step, we performed a comprehensive literature search using the search terms and search engines depicted in Table 1. Because we originally planned to examine the relationship between late responding and data quality (especially item nonresponse), we initially used additional search terms in the literature search. Unfortunately, it turned out that less than a third of the studies reported the data quality separately for late respondents. For further detail see Section 5.1.1.

To prevent a possible publication bias, we employed additional search strategies as follows: (1) we checked the reference list of key publications and (2) the published abstracts of several relevant conferences in the field of survey research (i.e., AAPOR, ESRA, GOR) and the ASA Online Proceedings of the Survey Research Methods Section from 2001 to 2020 for relevant abstracts.[2] An overview of our research strategy can be found in the PRISMA flow diagram (Fig. 1). The search strategy resulted in 53 full-text peer-reviewed manuscripts and 1 conference proceeding. The number of included studies however was 74 because 12 manuscripts included data from more than one independent sample (6 manuscripts reported data for 2 samples, 4 for 3, and 2 for 4), data from 74 independent study samples (n_{web} = 22; n_{mail} = 52) were included in the systematic review. As for 5 manuscripts, information on the number of late respondents was missing, it was not possible to calculate an effect size for these. Therefore, the sample for the meta-analytic procedures was smaller and comprised 50 manuscripts respectively 64 independent survey samples (n_{web} = 17; n_{mail} = 47).

[2] The additional search in the conferences abstracts and online proceedings resulted only in one additionally retrieved manuscript for our study sample. For 25 conference that passed our exclusion criteria, no corresponding publication could be identified. Because a similar outcome was expected for the following relevant conferences abstracts and proceedings no additional searches were conducted for International Total Survey Error Workshop (ITSEW), New Techniques and Technologies for Statistics (NTTS), International Methodology Symposium (IMS), International Conference on Computational Social Science (IC2S2), but resulted in no further relevant mansucript.

4.1.2 Coding Procedure

To include the most comprehensive and relevant survey characteristics, an iterative process was used to identify relevant codes. From the research literature, we consulted overviews summarizing moderator variables that are known to impact the response rate (e.g., Dillman, 2021; Groves & Peytcheva, 2008; Keusch, 2015). In addition, we amended potentially relevant moderator variables from the assessed manuscripts. For a better overview, we followed Groves and Peytcheva (2008). We coded publication-related characteristics of each study, such as publication date, publication journal, or field of research, in which the study was published. Additionally, we coded the characteristics of the target population and sample, such as the region where the study was conducted, whether the survey targeted experts or the general population or some sub-population, the total sample size, and the response rate. Furthermore, we also coded a number of survey design characteristics, such as the number of reminders sent, length of the survey, the field period, and whether the survey invitation was personalized (Joinson et al., 2007; Joinson & Reips, 2007) or incentives were promised. The outcome variables were the N and share of late respondents, for which we additionally coded the definition for late responding used in the study. In the data form that can be accessed via the link https://ehb.eu.qualtrics.com/jfe/form/SV_agdWokJe2bMIj4y, all covariate codes are listed. A to-

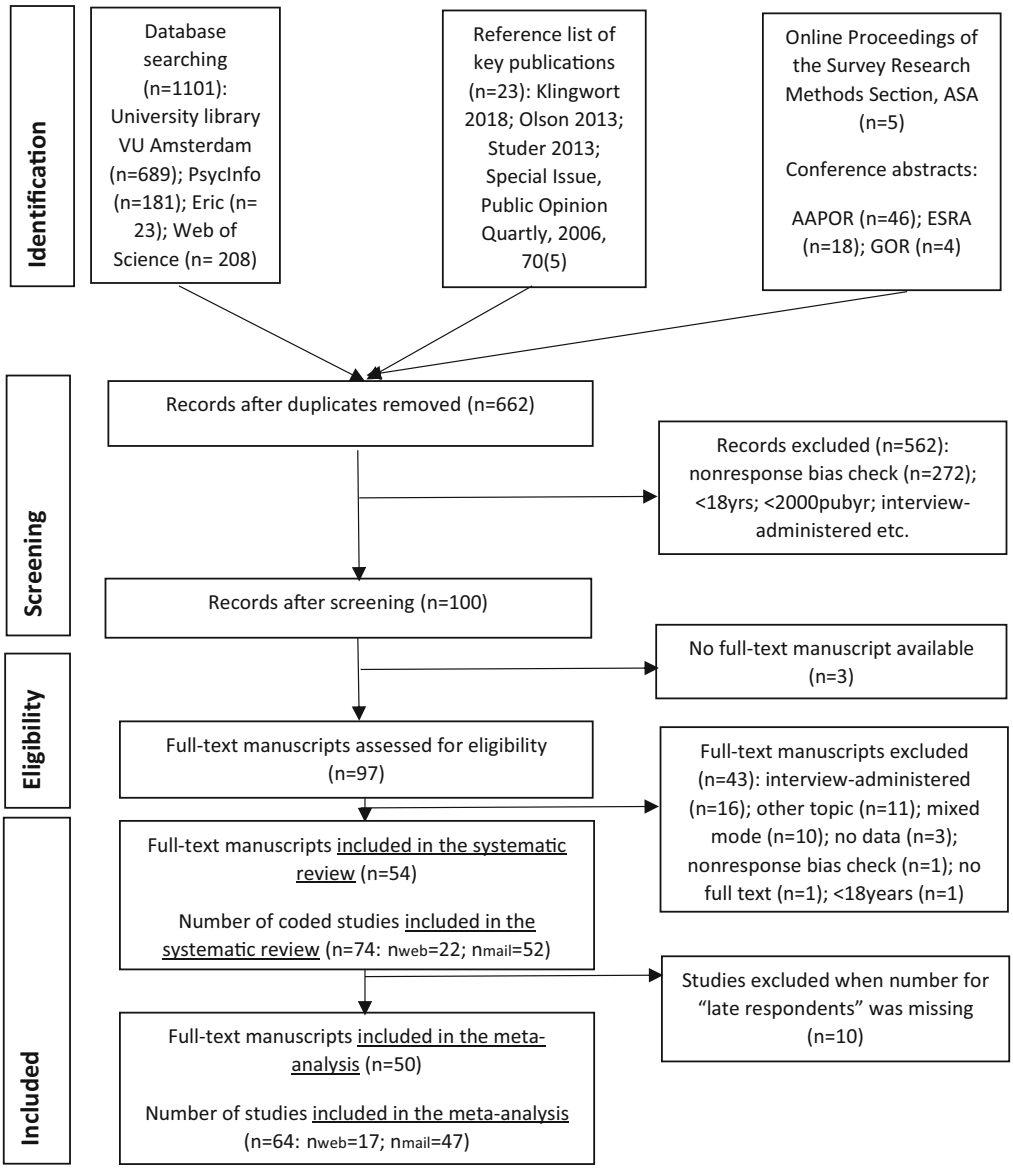

Fig. 1

PRISMA Flow Diagram for the Literature Search Process

tal of 46 different variables were coded per study sample. The coding was conducted by two independent coders. To ensure the quality of the coding, some standard procedures were followed throughout the process. i) All studies were reviewed independently by both coders; ii) After reviewing 10 studies each, results were compared; iii) Discrepancies (if any) were resolved through discussion; iv) Equal attention was given to each study; v) The process was thorough, rigorous, inclusive, and comprehensive; vi) The insights of both coders were fully acknowledged.

If information was missing for a category (e.g., the number of reminders), we tried to extract this information from additional publications or to calculate it from existing information (e.g., the number of late respondents). To answer the first question regarding common operationalizations of the phenomenon, the study authors' definitions of late responding was also included in the data set.

4.2 Meta-Analytic Procedures

To answer the second research question, we performed an overall meta-analysis and considered the two survey modes, web and mail, separately (Kaufmann et al., 2016). In our meta-analysis, we used the Hunter–Schmidt approach (see Schmidt & Hunter, 2014), which assumes a random ef-

fects model for the estimation. Using this approach, we consider not only within-study variance but also between-study variance (e.g., measurement error). To calculate the effect sizes, we took the total number of survey respondents and divided this by the number of survey respondents reported by the study authors as late respondents. In one case with insufficient data, we used the authors' definition of "late respondents" and calculated the share of late respondents from the reported response rate with and without late respondents (Friedman et al., 2003). To transform the observed proportions to a normal distribution, we applied a so-called Freeman-Tukey double arcsine transformed proportion (see Freeman & Tukey, 1950). The effect sizes were then weighted by the inverse-sampling variance and aggregated according to the Hunter–Schmidt meta-analytic approach.

Several strategies were employed to test for heterogeneity. First, we visually checked heterogeneity across studies using forest plots (Lewis & Clarke, 2001). Second, we examined heterogeneity between studies empirically by calculating the heterogeneity measures I^2 and Tau^2 (Higgins & Thompson, 2002). A low value for I^2 indicates low heterogeneity among studies. In contrast, a low Tau^2 indicates high heterogeneity. Additionally, we checked the robustness of our overall analysis and its dependency due to study outliers (Viechtbauer & Cheung, 2010), publication bias[3] (see the trim–fill approach suggested by Duval & Tweedie, 2000), or publication dependency. This robustness check was followed by a mode-specific analysis.

To answer the third research question, we considered the study characteristics as potential moderator variables (see below). We first checked the number of studies for each potential moderator variable in our database. Due to power issues, we only considered potential moderator variables if the sample of each mode included at least four studies. To avoid that results would be impacted by the weighting strategy of our meta-analysis (that is the number of participants), we also excluded the study characteristic "sample size". With all potential moderator variables, we ran a meta-analysis on both modes together, as well as for the two modes separately, to check for any mode dependency by overlapping confidence intervals. To check the robustness of our results, we also ran meta-regressions to examine the influence of a moderator variable (e.g., mode) on the proportion of late respondents (see Berkey et al., 1995). The analysis was carried out using R (version 4.0.5, R Core Team, 2020) and the metafor package (version 4.0.0, Viechtbauer, 2010).

[3] High-quality meta-analyses now use several strategies to prevent publication bias (e.g., asking authors in the field for unpublished data, including unpublished studies in the sample, see our literature search procedure) and also examine publication bias.

5 Results

5.1 Study Sample: Systematic Review

Although no restrictions were imposed on the type of literature and an additional search was conducted for conference papers, the 74 studies included in the final sample for the systematic review were published in peer-reviewed journals, except for one study, and included 45 different titles. Of the studies published between 2000 and 2021 in our sample, most of the mail survey studies (more than 70%) were published between 2000 and 2010. The first web survey study included was published in 2004; half of the web survey studies were not published before 2012. Not surprisingly, the mean sample size for the mail survey studies was 8,525, whereas that for the web survey studies was 38,814. However, the biggest proportion of studies (49% of the mail and 64% of the web survey studies) had a sample size falling between 1001 and 10,000, The sample sizes ranging from 170 (Mehlkop & Becker, 2015) to 135,000 (Friedman et al., 2003) for the mail survey studies and 115 (Parker et al., 2012) to 560,084 (Sigman et al., 2014) for the web survey studies. For more descriptive details of the study sample, see Table 4.

5.1.1 Subsample: Data Quality

A third of the studies (24 out of 74 studies) in 16 manuscripts reported one or more data quality indicators and only little than one fifth (17 out of 74 studies) reported a comparison between early and late respondents (10 studies for web and 7 studies for mail surveys, see Table A1 in the Appendix). The most frequent reported data quality indicator was item nonresponse (12 studies respectively 16%), followed by beak-off rate (4 studies respectively 5%) and the quality of open-end questions (e.g., nr. of words, nr. of themes) (3 studies in one manuscript). Other reported data quality indicators were motivated misreporting, nonsubstantive answers, and measurement equivalence.

Due to the small proportion of studies reporting on different type of data quality indicators for early and late respondents, any conclusions drawn must be considered with caution. In general, it can be concluded from Table A1 in the Appendix that the findings are mixed for mail as well as for web surveys. Mixed results seem the case not only between studies but also within studies, as late respondents react differently from early respondents to some experimental manipulations and not to others. Furthermore, differences are found across various data quality indicators. All in all, these finding suggest that late responding has merit to some extent also for web surveys.

Table 2

Overview of the Classification of Manuscripts into the Three Identified Late-Responding Definition Categories separately by Mode

		Mode	
Operationalization	Definition	Web	Mail
Threshold	1st reminder	Estelami (2015)[a]	Barclay et al. (2002); Bilodeau (2006); Dykema et al. (2015); Fejer (2006); Helasoja et al. (2002); Hox et al. (2012); Korkeila et al. (2001); Lie et al. (2019); Manjer et al. (2002); Menachemi et al. (2006); Nakai et al. (2005); Olowokure et al. (2004); Oremus and Wolfson (2004); Parashos et al. (2005); Randall (2015); Rashidian et al. (2008); Rueegg et al. (2017)
	2nd reminder	Giroux et al. (2020); Kypri et al. (2011); Sigman et al. (2014)	Friedman et al. (2003); Maclennan et al. (2012); Puleo et al. (2002)
	After personal contact	Parker et al. (2012)	Clendenning et al. (2013); Collins et al. (2000)
	Metric approach	Smyth et al. (2009)	Perneger et al. (2005); Vink and Boomsma (2008)
Extreme group	Before 1st vs. after 2nd reminder	Hansen et al. (2007); Klingwort et al. (2018); Rao and Pennington (2013)	Chen et al. (2003); Hazell et al. (2009); Lahaut et al. (2003); Maitland et al. (2017); Mallen et al. (2005); Meiklejohn et al. (2012)
	Before 1st vs. after a later reminder	–	Brehaut et al. (2006)
	Before 2nd vs. after 4th reminder	Jouriles et al. (2020)	–
	Before 1st reminder vs. after personal contact	Kypri et al. (2004)	Gasquet et al. (2001)
	Metric approach	Blank et al. (2009); Witry et al. (2021)	Kowall et al. (2010); Mott et al. (2001); Zachry et al. (2003)
Continuous	Response delay in days	Göritz and Stieger (2009); Gummer and Struminskaya (2020)	Gummer and Struminskaya (2020); Mehlkop and Becker (2015)
	Number of reminders	Bach et al. (2020)	Brøgger et al. (2003); Dykema et al. (2021)

[a] All studies except for Göritz and Stieger (2009), Gummer and Struminskaya (2020), Mehlkop and Becker (2015), and Zachary et al. (2003) are also included in the meta-analysis subsample.

Only by looking at it in a systematic way we can understand the phenomenon and is constituting factors. Moreover, it seems plausible to conclude that there is more to it than the simple passing of time and that it is therefore not a linear process, but rather that a pronounced motivational component is involved.

5.1.2 Subsample: Meta-Analysis

From our systematic review sample (see above) ten studies reported in four manuscripts had to be excluded for the meta-analytic sample (see also Table 2). They reported the results of five mail and five web survey studies each. Furthermore, they were all published in peer-reviewed journals, with publication years spread throughout the search period: 2003, 2009, 2015, and 2020. With the exception of the Mehlkop and Becker study (2015), for all excluded studies the sample size falls into the 1001-100,000 category. For the meta-analysis subsample, only one manuscript containing a comparison of data quality for early and late respondents had to be excluded (Göritz & Stieger, 2009). However, they reported results for four web survey studies.[4]

[4] One manuscript is mentioned under both the web and mail condition. This manuscript reports a study for both modes.

5.2 Operationalization of Late Responding

Having established that late responding is by no means a clearly defined phenomenon, the first step was to explore the range of definitions used in the literature and an attempt to provide a meaningful categorization. This then allowed a meaningful comparison between modes. Moreover, only by having such groundwork will it be possible to discuss the relationship between different definitions of late responding in a more systematic way.

Table 2 provides an overview of the three proposed categories, as well as the classification of the included manuscripts. Just over half of the studies (51%; n = 38) conceptualized late responding as a dichotomy with a *threshold*, below which respondents were classified as early respondents, while above that threshold respondents were classified as late respondents. The most used threshold is the first reminder. Considerably fewer authors have defined the second reminder as the threshold. Others distinguish between respondents who respond to one of the written survey requests (early respondents) and those who can only be persuaded to participate after personal contact (late respondents). A few operationalizations were not guided by a "natural" event within the design, such as the date of sending a reminder or switching the contact mode but favored a more "metric approach."

For example, Smyth et al. (2009) defined the median response date as the threshold that divided respondents into early and late respondents, whereas Vink and Boomsma's (2008) threshold criterion was the response rate reaching an asymptote after 30 days.

The second definition, also quite frequently used, defines early and late respondents as two *extreme groups*. More than one-third of the studies (35%; n = 26) used such a definition. Most of these authors compared respondents to the initial (e-)mail, with respondents not completing the survey before the second reminder. A few researchers have defined other extreme group cutoffs (see Table 2). Also, within this category, some metric approaches can be found. Mott et al. (2001), for example, defined the first quintile of their survey sample as early and the last quintile as late respondents. Kowall et al. (2010) compared the first with the last third of respondents, whereas Blank et al. (2009) compared the respondents of the initial mail to the number of respondents in the 90th percentile of the survey field period.

Studies in the third and smallest category (13%; n = 10) defined late responding as a *continuous process*, i.e., a continuous change rather than a dichotomous either-or grouping or or one of increasing stages of "partial cooperation." The biggest share in this category operationalized late responding as a response delay in days.

5.2.1 Mode Differences in the Operationalization of the Late-Responding Phenomenon

When looking at the modes separately, certain differences in preferences regarding the chosen definition become apparent (see Fig. 2). Two-thirds of the mail survey studies used a threshold definition, while the web survey studies used the threshold definition with similar frequency as the extreme group definition. The continuous definition was the least frequently used in both modes.

Most studies choose a reminder as a natural threshold or cutoff, assuming that the study participants would have been nonrespondents without this further contact attempt. Furthermore, while there were some differences between the modes regarding the preferred definition of the phenomenon, the proportion of late respondents was a very heterogeneous measure within both modes. Given this disparity in definitions, for the planned comparison of the proportion of late respondents between the two modes, we tried to identify a large enough subsample to generate a homogeneous operationalization of the phenomenon. Because additional information on the proportion of respondents after the first reminder was available for 24 studies that originally used a different late-response operationalization, we were able to form a homogeneous subsample with a total of 49 studies using the "threshold 1st reminder" definition.

5.3 Share of Late Respondents

5.3.1 Systematic Review

Of the 74 studies included in the systematic review, more than two-thirds were mail survey studies (n = 52, 70%). If one considers that the remaining 22 web survey studies were for the most part not published before the end of the first decade, it is not surprising that their proportion in the study sample is lower. On the other hand, considering how much the number of web survey studies has increased over the last 10 years, web survey studies seem to be under-

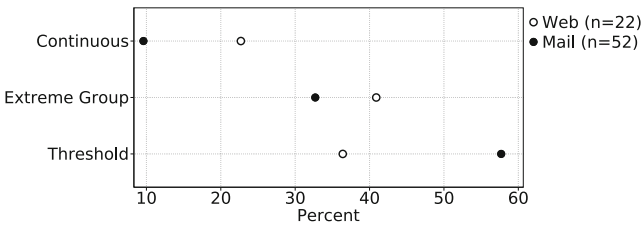

Fig. 2

Distribution of the Three Operationalizations of Late Responding over the Two Modes

Table 3

Meta-analysis of the Overall Proportion and Mode-Dependent Proportion of Late Respondents

						Heterogeneity Measures	
			N	% late	CI	I^2	Tau^2
Overall		Def. original sample	64	27	23–31	99.8	0.03
		Def. subsample	49	27	22–32	99.8	0.04
Mode	Def. original sample	Web	17	29	24–33	99.7	0.01
		Mail	47	27	22–32	99.8	0.04
	Def. subsample	Web	8	23	17–30	99.8	0.04
		Mail	41	27	22–33	99.8	0.04

represented in our sample (see, for example, Cornesse & Bosnjak, 2018).

5.3.2 Meta-analysis

Because in 10 studies data for calculating the effect sizes were not available, the database for the following meta-analysis comprised 64 studies, including 47 mail and 17 web studies.

Results for the Overall Sample. The meta-analysis of these 64 studies showed that across all studies, regardless of the phenomenon definition or mode, the overall weighted mean share for the late respondents was 27% (CI: 23%–31%). Table 3 presents an overview of the results and their robustness analyses (Figure A1 in the Appendix illustrates the forest plot). However, the data showed a high degree of heterogeneity. Because more than one study was reported in several manuscripts, which ranged from two to four studies in one manuscript, a dependence analysis was indicated. This two-level analysis confirmed our results (27%, CI: 22%–32%). There was also no outlier study (Figure A2 in the Appendix). Therefore, we performed a classical meta-analysis in the following.[5] Although running a publication bias estimation using the trim–fill approach showed one missing study, the overall results were nevertheless confirmed. Thus, we assume that our overall results are independent of any outliers, publication bias, or publication outlet dependency.

Our result was further confirmed, as the same analysis performed with the purpose-built subsample with a homogeneous definition of late responding using the first reminder as a threshold (as described in Section 5.1.2) showed the same result (see Table 3).

Mode-Separated Results. As a next step, we checked whether the proportion of late respondents differed between the two self-administered modes of interest. Again, we performed the first analysis with the original sample and a second analysis with the smaller, more homogeneous subsample. For both samples tested, we did not find significant mode differences for the share of late respondents. For the web survey mode, the weighted mean was 29% (CI: 24%–33%) for the original sample and 23% (CI: 17%–30%) for the subsample. For the mail survey mode, the weighted mean was 27% for both samples ($CI_{original}$: 22%–32% and $CI_{subsample}$: 22%–33%). A meta-regression analysis confirmed a lack of mode differences (see Appendix, Table A2, $p = 0.61$). We conclude that the mode is irrelevant to the share of late respondents (see the mode-specific forest plots showing the overlap of studies' confidence intervals between Figures A3 and A4 in the Appendix). Due to the fact that in both modes large heterogeneity is indicated (I^2 = 99.7–99.8, Tau^2 = 0.01–0.04, see Table 3), moderator variables may explain the heterogeneity within our data.

5.4 Moderator Analysis

As we assumed that the survey practices differ for the two modes, and to determine whether other characteristics of the target population and sample or the survey design influence the share of late respondents, we examined those key characteristics separately for the two modes and conducted several moderator analyses. We controled for heterogeneity in the following ways.

[5] We highlight that often meta-analyses are not seen as multilevel analyses, although a classical meta-analysis represents a null-model of a multilevel-analysis and hence is also a multilevel analysis (see also Pastor & Lazowski, 2018).

5.4.1 Systematic Review

Target Population and Sample Characteristics. The descriptive results of the target population and sample characteristics are shown in Table 4. Whereas the largest share of mail survey studies was conducted in Europe, the largest share of web survey studies was conducted in North America. If we look at the focus of the journals in which the included studies were published, two-thirds of the studies using mail surveys were published in the field of health, whereas methodological questions dominate for studies using web surveys. For the population studied, one more difference between the two modes became apparent: whereas the largest share of mail survey studies was done with the general population, the largest share of web survey studies surveyed students. Not surprisingly, we also found mode differences in the reported sample sizes. This was also reflected in the respective mode-specific mean sample size (as reported in section 5.1). Furthermore, although the response rate difference between the two modes is lower than the average 11% found in the meta-analysis conducted by Daikeler et al. (2019), as to be expected, the mean response rate for the web surveys (48%) was almost 7% lower than for the mail surveys (54%), ranging from 19% (Rashidian et al., 2008) to 87% (Parashos et al., 2005) for mail and 6% (Witry et al., 2021) to 91% (Gummer & Struminskaya, 2020) for web surveys.

Survey Design Characteristics. Table 5 depicts the detailed descriptive results reported in this section. In our sample, web surveys tend to employ more persuasive techniques (Göritz, 2006; Joinson & Reips, 2007; Reips & Franek, 2004) than mail surveys. This is the case for the use of *incentivization* (web = 64%; mail = 33%) as well as for *personalization* (web = 55%; mail = 21%). Surprisingly, for the use of reminders, mode differences were not prominent. Despite the slightly higher mean for the *number of reminders sent* in the web survey studies (M_{web} = 2.5), compared with the mail studies (M_{mail} = 2.1), the difference was not significant ($t(69) = 1.08$, $p > 0.05$). The *time interval in days* within which the first reminder was sent differed considerably for the two modes (M_{mail} = 22 vs. M_{web} = 6.7, $t(52) = 2.47$, $p < 0.05$). For the included web survey studies, on average, the first reminder was sent after one week, whereas for the mail survey studies, the first reminder was sent around three weeks after the first survey invitation on average. We also considered *survey length*, although for half of the mail survey studies (52%) and almost a quarter of the web survey studies (23%), information on survey length, whether in the form of the number of items or the average completion time, was missing. More than a fifth of the web survey studies were conducted with very short surveys (≤5 min or 27 items) compared to only one mail survey study. While the data for this variable are equal and complete for both modes, the proportion of surveys with short *field periods* (≤28 days) was larger for the web survey studies (web = 36%; mail = 10%).

5.4.2 Meta-analysis

As can be concluded from the systematic review, the included mail and web survey studies differed in several key study variables. Therefore, a moderator analysis can provide insights into the possible relevant dynamics to (de)motivate people to respond faster. As mentioned previously, for several reasons, the dataset for performing the moderator analyses was smaller than that for the systematic review. As continuous variables, such as response rate, number of reminders, or survey duration, were usually coded in four or five categories and information on some of the coded survey design variables was sometimes incomplete, several variable categories contained fewer than four studies. Accordingly, these variables were dichotomized in a further step so that more studies per category could be considered, and thus more studies could be included in the moderator analyses. In doing so, we had to accept that this would allow for less differentiated conclusions on possible moderators. Our final dataset, with the dichotomized variables, consisted of 16 potential moderator variables (Table 6). Of the total of 16 moderator variables included in the moderator analyses, 7 were associated with the target population and sample and 9 variables with the survey design.

Results for the Overall Sample. For the 16 moderator analyses performed, no significant moderator could be identified for the overall sample. The analysis resulted in a consistent share of late respondents, ranging from 22% for studies with a 0 to 14 days interval for the first reminder to 31% for studies conducted with experts (see Table 6).

An additionally conducted meta-regression identified one moderator for the overall sample. That is, studies conducted with the general population resulted in a higher late response rate than those with other more homogeneous target populations (e.g., patients, population-at-risk, students, health specialists) (Table A2 in the Appendix, rightmost columns).

Mode-Separated Results. For the moderator analysis for the web survey mode (Table 6, in the center left columns), the average proportion of late respondents ranged between 20% for web surveys targeting the general population and 41% for long surveys with a field period of 57 and more days. In the mail survey mode (Table 6, in the center right columns), the share of late respondents ranged from 17% for the mail

Table 4

Numbers and Percentages (%) of Possible Sample Characteristic Moderator Variables

	Mail				Web				Total			
	All		Moderator		All		Moderator		All		Moderator	
	N	%	n	%	n	%	n	%	n	%	n	%
Region												
North America	13	25	12	26	11	50	11	65	24	32	23	36
Europe	31	60	27	57	8	36	3	18	39	53	30	47
Australia/Oceania	7	14	7	15	2	9	2	12	9	12	9	14
South America	0	0	0	0	1	5	1	6	1	1	1	2
Asia	1	2	1	2	0	0	0	0	1	1	1	2
Field of publication												
Health	35	67	34	72	7	32	7	41	42	57	41	64
Methodology	7	14	6	13	12	55	7	41	19	26	13	20
Marketing/Business	0	0	0	0	1	5	1	6	1	1	1	2
Social Sciences	9	17	6	13	1	5	1	6	10	14	7	11
Education	0	0	0	0	1	5	1	6	1	1	1	2
Psychology	1	2	1	2	0	0	0	0	1	1	1	2
Expertise												
Experts yes	12	23	12	26	4	18	4	24	16	22	16	25
Sample												
General Population +	31	60	27	57	5	23	4	24	36	49	31	48
Patients/Population at risk	6	12	6	13	0	0	0	0	6	8	6	9
Employees	0	0	0	0	1	5	1	6	1	1	1	2
Students	1	2	1	2	13	59	9	53	14	19	10	16
Health Specialists	13	25	12	26	3	14	3	18	16	22	15	23
Others	1	2	1	2	0	0	0	0	1	1	1	2
Response rate (%)												
0–25	1	2	1	2	5	23	4	24	6	8	5	8
26–50	22	42	19	40	7	32	5	29	29	39	24	38
51–75	19	37	18	38	7	32	6	35	26	35	24	38
76–100	7	14	6	13	3	14	2	12	10	14	8	13
Missing	3	6	3	6	0	0	0	0	3	4	3	5
N	52		47		22		17		74		64	

The "moderator" columns report the number of studies that were included in the moderator analyses for each moderator variable. Only moderator variables with four or more studies were considered for the moderator analyses.

surveys conducted in North America to 30% for the mail surveys targeting experts.

As for four variables the respective mode-specific confidence intervals for the proportion of late respondents did not overlap, four potential mode-specific moderators were identified (Table 6, in the leftmost column): i.e., when the studies were conducted in North America, when the studies were published in a methodological journal, when the field period of the studies was 57 days or longer, and when the survey length was (very) short, the share of late respondents in the web survey studies was potentially higher compared to that in mail survey studies.

To check the robustness of our results, additional meta-regressions were performed. Because the multiple meta-regression analyses failed to uncover any potential moderator variable consistently across both modes, we concluded that the sample of potential moderator variables examined could not reliably explain the proportion of late respondents.

6 Discussion

The results presented in this paper add several more pieces to the puzzle of differences between survey modes and pro-

Table 5

Numbers and Percentages (%) of Possible Survey Design Characteristic Moderator Variables

Survey design characteristics	Mail				Web				Total			
	All		Moderator		All		Moderator		All		Moderator	
	n	%	n	%	n	%	n	%	n	%	n	%
Panel												
Panel yes	4	8	3	6	7	32	3	18	11	15	6	9
Number of reminders												
0	4	8	0	0	4	18	0	0	8	11	0	0
1	9	17	9	19	1	5	1	6	10	14	10	16
2 +	22	42	21	45	8	36	7	41	30	41	28	44
3	14	27	14	30	1	5	1	6	15	20	15	23
>3	3	6	3	6	5	23	5	29	8	11	8	13
Missing	0	0	0	0	3	14	3	18	3	4	3	5
Interval to first reminder in days												
0	0	0	0	0	4	18	0	0	4	5	0	0
≤7	12	23	11	23	8	36	8	47	20	27	19	30
8–14	10	19	10	21	4	18	4	24	14	19	14	22
15–21	5	10	5	11	0	0	0	0	5	7	5	8
22–28	8	15	8	17	0	0	0	0	8	11	8	13
>28	5	10	5	11	0	0	0	0	5	7	5	8
Missing	12	23	8	17	6	27	5	29	18	24	13	20
Incentivization												
Incentive yes	17	33	14	30	14	64	10	59	31	42	24	38
Personalization												
Personalization yes	11	21	11	23	12	55	8	47	23	31	19	30
Survey length												
Very short (≤5 min, ≤27 items)	1	2	1	2	5	23	4	24	6	8	5	8
Short (>5–15 min, 28–50 items)	9	17	9	19	4	18	4	24	13	18	13	20
Medium (>15–25 min, 51–100 items)	9	17	9	19	6	27	3	18	15	20	12	19
Long (>25 min, >100 items)	6	12	6	13	2	9	2	12	8	11	8	13
Missing	27	52	22	47	5	23	4	24	32	43	26	41
Field period in days												
≤28	5	10	1	2	8	36	4	24	13	18	5	8
29–56	24	46	23	49	6	27	6	35	30	41	29	45
57–84	6	12	6	13	2	9	1	6	8	11	7	11
>84	12	23	12	26	4	18	4	24	16	22	16	25
Missing	5	10	5	11	2	9	2	12	7	10	7	11
N	52		47		22		17		74		64	

The "moderator" columns report the number of studies that were included in the moderator analyses for each moderator variable. Only moderator variables with four or more studies were considered for the moderator analyses.

Table 6

Meta-analysis of Potential Moderator Variables on the Target Population and Sample and the Survey Design (Including Dichotomized Variables)

	N				Web				Mail				Total			
	Web	Mail	Total	Total Sample (%)			Heterogeneity measure				Heterogeneity measure				Heterogeneity measure	
Variable	17	47	64	100.00	Late (%)	Interval	I^2	Tau^2	Late (%)	Interval	I^2	Tau^2	Late (%)	Interval	I^2	Tau^2
Target Population & Sample																
Experts: yes	4	12	16	25.00	35	30–40	74.3	0.00	30[b]	22–39	98.6	0.02	31[b]	25–38	98.4	0.02
Sample: General Population	4	27	31	48.44	20[b]	14–26	99.5	0.01	25	18–32	99.9	0.05	23	18–30	99.8	0.04
Response rate: <= 50%	9	20	29	43.75	23	19–26	99.2	0.00	26	18–36	99.9	0.06	25	19–30	99.8	0.02
Response rate: > 50%	8	24	32	50.00	32	24–41	99.5	0.02	26	21–32	99.6	0.03	27	22–32	99.6	0.03
Country: North America[a]	11	12	23	35.94	34	28–39	99.6	0.00	17[b]	11–22	99.5	0.02	24	18–30	99.9	0.03
Field: Health	7	34	41	64.06	22	17–28	98.8	0.01	26	20–33	99.8	0.04	25	20–31	99.8	0.03
Field: Method[a]	7	6	13	20.31	36	33–40	98.5	0.00	23	17–30	98.8	0.01	30	23–38	99.8	0.02
Survey design																
Field period: 0 to 56 days	10	24	34	53.13	24	18–30	99.8	0.01	26	20–32	99.7	0.03	25	20–30	99.8	0.03
Field period: 57 and more days[a]	5	18	23	35.94	41[b]	34–48	94.5	0.00	22	18–27	99.1	0.01	26	21–31	99.2	0.02
Interval to first reminder: 0 to 14 days	12	21	33	51.56	24	19–29	99.7	0.10	21	16–26	99.5	0.02	22[b]	17–27	99.8	0.03
Survey length: (very) short[a]	7	10	17	26.56	36	26–48	99.6	0.02	21	16–25	96.3	0.01	27	21–33	99.4	0.01
Survey lenght: medium and long	5	15	20	31.25	29	26–32	97.9	0.00	26	20–33	99.5	0.02	27	24–30	99.4	0.01
Reminders: 1 to 2	8	30	38	59.38	26	20–34	99.6	0.01	28	21–35	99.8	0.05	28	20–34	99.8	0.04
Reminders: 3 and more	6	17	23	35.94	30	17–45	99.5	0.04	24	15–33	99.6	0.05	26	18–33	99.6	0.04
Personalized: yes	8	11	19	29.69	29	21–39	99.4	0.02	25	14–37	99.8	0.05	27	19–35	99.7	0.05
Incentive: yes	10	14	24	37.50	30	24–36	99.0	0.01	25	14–37	99.8	0.06	27	20–35	99.7	0.05

[a] Potential mode-specific moderator variables with nonoverlapping confidence intervals for the two modes
[b] Min and max values

vide aggregate information on late responding in web surveys for the first time. Our research has implications for future research on late responding, which we will discuss in Section 6.3 below.

6.1 Summary

To the best of our knowledge, this paper offers the first systematic attempt to describe and categorize the wide variety of definitions of late responding used in the research literature. Our findings confirm that the definitions used vary considerably within and across both modes. While half of the studies use a dichotomous definition identifying a threshold that separates early and late respondents, the reviewed literature offers an either–or approach, an increasing level of "partial cooperation" approach, or a linear approach of "partial cooperation." Although some deductions and findings point to a motivational component, it remains ambiguous what constitutes the phenomenon of late responding.

With the meta-analytic approach chosen, we identified the overall average share of late respondents to be slightly more than a quarter of the total sample (around 27%). This result was confirmed by an additional analysis using a subsample with a homogeneous phenomenon definition. Nevertheless, despite identified differences in operationalization preferences and considerable differences in the target population, sample, and survey design practices between the mail and web survey studies, no mode effect was detected. One reason for this could be that, not only between but also within each of the two-mode subsamples, survey design practices vary considerably. Moreover, the average mode difference for one of the key variables—that is, the number of reminders—was rather small. It seems that the commonly applied practices of the web survey mode, such as shorter reminder intervals and field times as well as the more frequent use of refusal aversion techniques lead to a similar result as for mail surveys. Nevertheless, if this generous quarter provides worse survey data, that is a relevant proportion for web and mail surveys a like, and we definitively should know more about it.

With the moderator analysis, the paper addresses another little researched aspect of the late-responding phenomenon. However, across the two modes, none of the tested potential moderator variables were confirmed consistently, and the identified three web survey-specific moderator variables must be interpreted with caution for various reasons. Our results indicate that moderator variables influence late responding differently in the two modes. However, the identified heterogeneity between and within the two modes plays into this as well. Given these results and considering the possible interdependencies of different moderator variables, further research into what moderators contribute to late responding is needed.

6.2 Limitations

There are however some restrictions that limit the generalizability of the findings. One main problem was the missing data at different levels, which affected the robustness of the results in a variety of ways. Due to missing data about the share of late respondents, ten studies had to be excluded from the meta-analytical analyses because no effect size could be calculated. Additionally, information on the relevant survey design variables in the studies was sometimes incomplete. In particular, information on the length of the survey or the length of the field period as well as the information whether a study was a panel study or not, was missing for a substantial number of studies. Given these limitations in the dataset, which also led to a broad categorization of moderator variables into mostly only two levels, we must assume that relevant moderators or moderator categories may go undetected. Furthermore, the potential mode-specific moderator variables that were identified should be interpreted with caution. Last but not least, due to the lack of detailed information on data quality for late respondents, especially item nonresponse, no conclusions could be drawn on possible relationships between the different definitions of late responding and data quality. However, we hope that with the current Open Science movement, in the future, more data will be available for rerunning our analyses to overcome any missing data problems we faced in our meta-analysis (see also Kaufmann et al., 2016).

Another limitation is that probable selectivity of the study sample must be assumed because the keywords for the literature search were rather narrowly defined, and the final sample consisted almost exclusively of peer-reviewed studies. Moreover, the subsample for the web surveys was rather small. Nevertheless, the additional measures taken to control for these issues, such as an additional search for unpublished conference and methods papers, multiple robustness checks for outliers, publication bias, and publication dependencies, revealed no evidence of missing publications.

Finally, as some moderator variables may not be independent from each other, more complex analyses are needed to reveal possible patterns of moderators. This further strengthens the notion that participants' response behavior is influenced by more than one survey characteristic.

6.3 Implications for Survey Research and Practice

Our research findings not only suggest that late responding is a substantive phenomenon for web and mail surveys alike, but they also reveal various important research gaps. First, more systematic theoretical and empirical groundwork is needed to capture the phenomenon more precisely. In doing so, relationships to data quality, survey outcomes, and sample characteristics need to be more systematically considered and tested for variations in survey (design) characteristics. Given a large number of possible survey design variations and taking into account the interdependence of different design decisions, further studies, especially web surveys, are needed that carefully and comprehensively report all relevant sample and design characteristics. Another promising line of inquiry might be to examine whether different reasons for late responding lead to different effects on different indicators of data quality. From there, one could investigate whether, for different modes, other people respond with delay and whether the patterns of reasons vary and thus affect data quality differently.

As for survey practice, these findings underscore that survey practitioners need not worry that the proportion of late respondents will increase if they use a web survey instead of a mail survey. Furthermore, based on the most commonly identified practice, it is recommended to define "late respondents" as respondents completing the survey only after the first reminder is sent. From this baseline, it would then be meaningful to look at other commonly used definitions and systematically examine their effects on data quality and relevant study outcomes. Our meta-analytic findings suggest that a 21–31% proportion of late respondents in a mail or web survey can be considered "normal," whereas a considerably lower or higher share can be taken as a motivation for conducting a more in-depth analysis of the possible reasons to better understand the underlying dynamics.

Conflict of interest We have no conflicts of interest to disclose.

Preliminary findings were presented at the General Online Research Conference (GOR) in 2021.

The references of the studies included in the meta-analysis are given in Table 2.

References

Bach, R.L., Eckman, S., & Daikeler, J. (2020). Misreporting among reluctant respondents. *Journal of Survey Statistics and Methodology*, *8*(3), 566–588. https://doi.org/10.1093/jssam/smz013.

Barclay, S., Todd, C., Finlay, I., Grande, G., & Wyatt, P. (2002). Not another questionnaire! Maximizing the response rate, predicting non-response and assessing non-response bias in postal questionnaire studies of GPs. *Family Practice*, *19*(1), 105–111. https://doi.org/10.1093/fampra/19.1.105.

Berkey, C.S., Hoaglin, D.C., Mosteller, F., & Colditz, G.A. (1995). A random-effects regression model for meta-analysis. *Statistics in Medicine*, *14*(4), 395–411.

Bilodeau, A. (2006). Non-response error versus measurement error: a dilemma when using mail questionnaires for election studies. *Australian Journal of Political Science*, *41*(1), 107–118. https://doi.org/10.1080/10361140500507310.

Blank, D., Hohgraefe Neto, G., Grando, E., Siqueira, P.Z., Lunkes, R.P., Pietrobeli, J.L., Marzola, N.R., & Goldani, M.Z. (2009). A web-based survey on students' conceptions of "accident". *Informatics for Health and Social Care*, *34*(4), 189–208. https://doi.org/10.3109/17538150903358057.

Brehaut, J.C., Graham, I.D., Visentin, L., & Stiell, I.G. (2006). Print format and sender recognition were related to survey completion rate. *Journal of Clinical Epidemiology*, *59*(6), 635–641. https://doi.org/10.1016/j.jclinepi.2005.04.012.

Brøgger, J., Bakke, P., Eide, G.E., & Gulsvik, A. (2003). Contribution of follow-up of nonresponders to prevalence and risk estimates: A Norwegian respiratory health survey. *American Journal of Epidemiology*, *157*(6), 558–566. https://doi.org/10.1093/aje/kwg003.

Callegaro, M., Manfreda, K.L., & Vehovar, V. (2015). Web survey methodology. Sage.

Čehovin, G., Bosnjak, M., & Lozar Manfreda, K. (2022). Item nonresponse in web versus other survey modes: a systematic review and meta-analysis. *Social Science Computer Review*. https://doi.org/10.1177/08944393211056229.

Chen, R., Wei, L., & Syme, P.D. (2003). Comparison of early and delayed respondents to a postal health survey: a questionnaire study of personality traits and neuropsychological symptoms. *European Journal of Epidemiology*, *18*(3), 195–202. https://doi.org/10.1023/A:1023393231234.

Clendenning, G., Field, D.R., & Jensen, D. (2013). A survey of seasonal and permanent landowners in Wisconsin's Northwoods: following Dillman and then some. *Society and Natural Resources*, *26*(2), 226–237. https://doi.org/10.1080/08941920.2013.767730.

Collins, R.L., Ellickson, P.L., Hays, R.D., & Mccaffrey, D.F. (2000). Effects of incentive size and timing on response rates to a follow-up wave of a longitudinal mailed survey. *Evaluation Review*, *24*(4), 347–363. https://doi.org/10.1177/0193841X0002400401.

Cornesse, C., & Bosnjak, M. (2018). Is there an association between survey characteristics and representativeness? A meta-analysis. *Survey Research Methods*, *12*(1), 1–13. https://doi.org/10.18148/srm/2018.v12i1.7205.

Crews, T.B., & Curtis, D.F. (2011). Online course evaluations. Faculty perspective and strategies for improved response rates. *Assessment & Evaluation in Higher Education*, *36*(7), 865–878. https://doi.org/10.1080/02602938.2010.493970.

Dahlhamer, J.M., Simile, C.M., & Taylor, B. (2008). Do you really mean what you say? Doorstep concerns and data quality in the National Health Interview Survey (NHIS). In *Proceedings of the Joint Statistical Meetings of the American Statistical Association* (pp. 1484–1491).

Daikeler, J., Bošnjak, 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.

Dickersin, K. (2005). Publication bias in meta-analysis: prevention, assessment and adjustments. In H.R. Rothstein, A.J. Sutton & M. Borenstein (Eds.), *Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm* (pp. 11–33). Wiley.

Dillman, D.A. (2021). 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. https://doi.org/10.1007/978-3-030-47256-6.

Dillman, D.A., Reips, U.-D., & Matzat, U. (2010). Advice in surveying the general public over the Internet. *International Journal of Internet Science*, *5*, 1–4.

Dillman, D.A., Smyth, J.D., & Christian, L.M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method* (4th edn.). Wiley.

Dodou, D., & De Winter, J.C.F. (2014). Social desirability is the same in offline, online, and paper surveys: a meta-analysis. *Computers in Human Behavior*, *36*, 487–495. https://doi.org/10.1016/j.chb.2014.04.005.

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463.

Dykema, J., Jaques, K., Cyffka, K., Assad, N., Hammers, R.G., Elver, K., Malecki, K.C., & Stevenson, J. (2015). Effects of sequential prepaid incentives and envelope messaging in mail surveys. *Public Opinion Quarterly*, *79*(4), 906–931. https://doi.org/10.1093/poq/nfv041.

Dykema, J., Stevenson, J., Assad, N., Kniss, C., & Taylor, C.A. (2021). Effects of sequential prepaid incentives on response rates, data quality, sample representativeness, and costs in a mail survey of physicians. *Evaluation & the Health Professions*, *44*(3), 235–244. https://doi.org/10.1177/0163278720958186.

Emberton, M., & Black, N. (1995). Impact of non-response and of late-response by patients in a multi-centre surgical outcome audit. *International Journal for Quality in Health Care*, *7*(1), 47–55. https://doi.org/10.1093/intqhc/7.1.47.

Estelami, H. (2015). The effects of survey timing on student evaluation of teaching measures obtained using online surveys. *Journal of Marketing Education*, *37*(1), 54–64. https://doi.org/10.1177/0273475314552324.

Fejer, R., Hartvigsen, J., Ohm Kyvik, K., Jordan, A., Christensen, H.W., & Høilund-Carlsen, P.F. (2006). The Funen Neck and Chest Pain Study: Analysing non-response bias by using national vital statistic data. *European Journal of Epidemiology*, *21*(3), 171–180.

Freeman, M.F., & Tukey, J.W. (1950). Transformations related to the angular and the square root. *Annals of Mathematical Statistics*, *21*(4), 607–611. https://doi.org/10.1214/aoms/1177729756.

Friedman, E., Clusen, N., & Hartzell, M. (2003). Better late? Characteristics of late respondents to a health care survey. In *Proceedings of the Survey Research Methods Section of American Statistical Association* (pp. 992–998).

Gasquet, I., Falissard, B., & Ravaud, P. (2001). Impact of reminders and method of questionnaire distribution on patient response to mail-back satisfaction survey. *Journal of Clinical Epidemiology*, *54*(11), 1174–1180. https://doi.org/10.1016/S0895-4356(01)00387-0.

Giroux, S.A., Gesselman, A.N., Garcia, J.R., Luetke, M., & Rosenberg, M. (2020). The magnitude and potential impact of missing data in a sexual violence campus climate survey. *Journal of American College Health*, *68*(5), 468–476. https://doi.org/10.1080/07448481.2019.1577865.

Gnambs, T., & Kaspar, K. (2017). Socially desirable responding in web-based questionnaires: a meta-analytic review of the candor hypothesis. *Assessment*, *24*(6), 746–762.

Goos, M., & Salomons, A. (2017). Measuring teaching quality in higher education: assessing selection bias in course evaluations. *Research in Higher Education*, *58*(4), 341–364. https://doi.org/10.1007/s11162-016-9429-8.

Göritz, A.S. (2006). Incentives in web studies: methodological issues and a review. *International Journal*

of Internet Science, *1*(1), 58–70. http://www.ijis.net/ijis1_1/ijis1_1_goeritz.pdf.

Göritz, A.S., & Stieger, S. (2009). The impact of the field time on response, retention, and response completeness in list-based Web surveys. *International Journal of Human Computer Studies*, *67*(4), 342–348. https://doi.org/10.1016/j.ijhcs.2008.10.002.

Groves, R.M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: a meta-analysis. *Public Opinion Quarterly*, *72*(2), 167–189. https://doi.org/10.1093/poq/nfn011.

Gummer, T., & Struminskaya, B. (2020). Early and late participation during the field period: response timing in a mixed-mode probability-based panel survey. *Sociological Methods and Research*. https://doi.org/10.1177/0049124120914921.

Hansen, R.A., Henley, A.C., Brouwer, E.S., Oraefo, A.N., & Roth, M.T. (2007). Geographic information system mapping as a tool to assess nonresponse bias in survey research. *Research in Social and Administrative Pharmacy*, *3*(3), 249–264. https://doi.org/10.1016/j.sapharm.2006.10.001.

Hazell, M., Morris, J., Linehan, M., Frank, P., & Frank, T. (2009). Factors influencing the response to postal questionnaire surveys about respiratory symptoms. *Primary Care Respiratory Journal*, *18*(3), 165–170. https://doi.org/10.3132/pcrj.2009.00001.

Helasoja, V., Prättälä, R., Dregval, L., Pudule, I., & Kasmel, A. (2002). Late response and item nonresponse in the Finbalt Health Monitor Survey. *European Journal of Public Health*, *12*(2), 117–123. https://doi.org/10.1093/eurpub/12.2.117.

Higgins, J.P.T., & Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186.

Hox, J.J., de Leeuw, E., & Chang, H.T. (2012). Nonresponse versus measurement error: are reluctant respondents worth pursuing? *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *113*(1), 5–19. https://doi.org/10.1177/0759106311426987.

Joinson, A.N., & Reips, U.-D. (2007). Personalized salutation, power of sender and response rates to web-based surveys. *Computers in Human Behavior*, *23*(3), 1372–1383. https://doi.org/10.1016/j.chb.2004.12.011.

Joinson, A.N., Woodley, A., & Reips, U.-D. (2007). Personalization, authentication and self-disclosure in self-administered Internet surveys. *Computers in Human Behavior*, *23*(1), 275–285. https://doi.org/10.1016/j.chb.2004.10.012.

Jouriles, E.N., Nguyen, J., Krauss, A., Stokes, S.L., & McDonald, R. (2020). Prevalence of sexual victimization among female and male college students: A methodological note with data. *Journal of Interpersonal Violence*. https://doi.org/10.1177/0886260520978198.

Kaminska, O., McCutcheon, A.L., & Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, *74*(5), 956–984. https://doi.org/10.1093/poq/nfq062.

Kaufmann, E., & Reips, U.-D. (2008). *Internet-basierte Messung sozialer Erwünschtheit: Theoretische Grundlagen und experimentelle Untersuchung [Internet-based measurement of social desirability: Theoretical foundations and experimental investigation]*. VDM Verlag Dr. Müller.

Kaufmann, E., Reips, U.-D., & Merki, K.M. (2016). Avoiding methodological biases in meta-analysis. *Zeitschrift für Psychologie/Journal of Psychology*, *224*(3), 157–167. https://doi.org/10.1027/2151-2604/a000251.

Keusch, F. (2015). Why do people participate in web surveys? Applying survey participation theory to Internet survey data collection. *Management Review Quarterly*, *65*(2), 183–216.

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. https://doi.org/10.3233/SJI-170421.

Korkeila, K., Suominen, S., Ahvenainen, J., Ojanlatva, A., Rautava, P., Helenius, H., & Koskenvuo, M. (2001). Non-response and related factors in a nationwide health survey. *European Journal of Epidemiology*, *17*(11), 991–999. https://doi.org/10.1023/A:1020016922473.

Kowall, B., Breckenkamp, J., Heyer, K., & Berg-Beckhoff, G. (2010). German wide cross sectional survey on health impacts of electromagnetic fields in the view of general practitioners. *International Journal of Public Health*, *55*(5), 507–512. https://doi.org/10.1007/s00038-009-0110-2.

Kreuter, F., Müller, G., & Trappmann, M. (2014). A note on mechanisms leading to lower data quality of late or reluctant respondents. *Sociological Methods & Research*, *43*(3), 452–464. https://doi.org/10.1177/0049124113508094.

Kypri, K., Stephenson, S., & Langley, J. (2004). Assessment of nonresponse bias in an internet survey of alcohol use. *Alcoholism: Clinical and Experimental Research*, *28*(4), 630–634. https://doi.org/10.1097/01.ALC.0000121654.99277.26.

Kypri, K., Samaranayaka, A., Connor, J., Langley, J.D., & Maclennan, B. (2011). Non-response bias in a web-based health behaviour survey of New Zealand tertiary students. *Preventive Medicine*, *53*(4–5), 274–277. https://doi.org/10.1016/j.ypmed.2011.07.017.

Lahaut, V.M.H.C.J., Jansen, H.A.M., Van De Mheen, D., Garretsen, H.F.L., Verdurmen, J.E.E., & Van Dijk, A. (2003). Estimating non-response bias in a survey on alcohol consumption: Comparison of response waves. *Alcohol and Alcoholism*, *38*(2), 128–134. https://doi.org/10.1093/alcalc/agg044.

Lewis, S., & Clarke, M. (2001). Forest plots: Trying to see the wood and the trees. *British Medical Journal*, *322*(7300), 1479–1480.

Lewis, T., Gorsak, M., & Yount, N. (2019). An automated refusal conversion strategy for web surveys. *Field Methods*, *31*(4), 309–327. https://doi.org/10.1177/1525822X19871016.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P.A., Clarke, M., Devereaux, P.J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine*. https://doi.org/10.1371/journal.pmed.1000100.

Lie, H.C., Rueegg, C.S., Fosså, S.D., Loge, J.H., Ruud, E., & Kiserud, C.E. (2019). Limited evidence of non-response bias despite modest response rate in a nationwide survey of long-term cancer survivors—Results from the NOR-CAYACS study. *Journal of Cancer Survivorship*, *13*(3), 353–363. https://doi.org/10.1007/s11764-019-00757-x.

Maclennan, B., Kypri, K., Langley, J., & Room, R. (2012). Non-response bias in a community survey of drinking, alcohol-related experiences and public opinion on alcohol policy. *Drug and Alcohol Dependence*, *126*, 189–194. https://doi.org/10.1016/j.drugalcdep.2012.05.014.

Maitland, A., Lin, A., Cantor, D., Jones, M., Moser, R.P., Hesse, B.W., Davis, T., & Blake, K.D. (2017). A nonresponse bias analysis of the Health Information National Trends Survey (HINTS). *Journal of Health Communication*, *22*(7), 545–553. https://doi.org/10.1080/10810730.2017.1324539.

Mallen, C., Peat, G., Thomas, E., & Croft, P. (2005). Severely disabling chronic pain in young adults: prevalence from a population-based postal survey in North Staffordshire. *BMC Musculoskeletal Disorders*, *6*(1), 42. https://doi.org/10.1186/1471-2474-6-42.

Manjer, J., Elmståhl, S., Janzon, L., & Berglund, G. (2002). Invitation to a population-based cohort study: differences between subjects recruited using various strategies. *Scandinavian Journal of Public Health*, *30*(2), 103–112. https://doi.org/10.1080/14034940210133771.

Mehlkop, G., & Becker, R. (2015). Wirkungen monetärer Anreize auf Antworten zur eigenen Delinquenz bei postalischer Befragung [Effects of monetary incentives on self-reports on delinquency in a mail survey]. *Zeitschrift Für Soziologie*, *44*(3), 215–228. https://doi.org/10.1515/zfsoz-2015-0305.

Meiklejohn, J., Connor, J., & Kypri, K. (2012). The effect of low survey response rates on estimates of alcohol consumption in a general population survey. *PLoS ONE*, *7*(4), 1–6. https://doi.org/10.1371/journal.pone.0035527.

Menachemi, N., Hikmet, N., Stutzman, M., & Brooks, R.G. (2006). Investigating response bias in an information technology survey of physicians. *Journal of Medical Systems*, *30*(4), 277–282. https://doi.org/10.1007/s10916-005-9009-5.

Morrison, R. (2011). A comparison of online versus traditional student end-of-course critiques in resident courses. *Assessment & Evaluation in Higher Education*, *36*(6), 627–641. https://doi.org/10.1080/02602931003632399.

Mott, D.A., Pederson, C.A., Doucette, W.R., Gaither, C.A., & Schommer, J.C. (2001). A national survey of U.S. pharmacists in 2000: assessing nonresponse bias of a survey methodology. *AAPS PharmSci*, *3*(4), 76–86. https://doi.org/10.1208/ps030433.

Nakai, Y., Milgrom, P., Yoshida, T., Ishihara, C., & Shimono, T. (2005). Evaluation of the total design method in a survey of Japanese dentists. *BMC Medical Research Methodology*, *5*(1), 27. https://doi.org/10.1186/1471-2288-5-27.

Olowokure, B., Caswell, M., & Duggal, H.V. (2004). Response patterns to a postal survey using a cervical screening register as the sampling frame. *Public Health*, *118*(7), 508–512. https://doi.org/10.1016/j.puhe.2003.12.013.

Olson, K. (2013). Do non-response follow-ups improve or reduce data quality? A review of the existing literature. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(1), 129–145. https://doi.org/10.1111/j.1467-985X.2012.01042.x.

Olson, K., Lepkowski, J.M., & Garabrant, D.H. (2010). An experimental examination of the content of persuasion letters on nonresponse rates and survey estimates in a nonresponse follow-up study. *Survey Research Methods*, *5*(1), 21–26.

Oremus, M., & Wolfson, C. (2004). Female specialists were more likely to respond to a postal questionnaire about drug treatments for Alzheimer disease. *Journal of Clinical Epidemiology*, *57*(6), 620–623. https://doi.org/10.1016/j.jclinepi.2003.11.008.

Parashos, P., Morgan, M.V., & Messer, H.H. (2005). Response rate and nonresponse bias in a questionnaire survey of dentists. *Community Dentistry and Oral Epidemiology*, *33*(1), 9–16. https://doi.org/10.1111/j.1600-0528.2004.00181.x.

Parker, M.J., Manan, A., & Urbanski, S. (2012). Prospective evaluation of direct approach with a tablet device as a strategy to enhance survey study participant response rate. *BMC Research Notes*, *5*(1), 1–6. https://doi.org/10.1186/1756-0500-5-605.

Pastor, D.A., & Lazowski, R.A. (2018). On the multilevel nature of meta-analysis: a tutorial, comparison of software programs, and discussion of analytic choices. *Multivariate Behavioral Research*, *53*(1), 74–89. https://doi.org/10.1080/00273171.2017.1365684.

Perneger, T.V., Chamot, E., & Bovier, P.A. (2005). Nonresponse bias in a survey of patient perceptions of hospital care. *Medical Care*, *43*(4), 374–380. https://doi.org/10.1097/01.mlr.0000156856.36901.40.

Peytchev, A., & Hill, C.A. (2010). Experiments in mobile web survey design: similarities to other modes and unique considerations. *Social Science Computer Review*, *28*(3), 319–335.

Puleo, E., Zapka, J., White, M.J., Mouchawar, J., Somkin, C., & Taplin, S. (2002). Caffeine, cajoling, and other strategies to maximize clinician survey response rates. *Evaluation & the Health Professions*, *25*(2), 169–184. https://doi.org/10.1177/016787020250020003.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/.

Randall, K.J. (2015). Early and late responders to questionnaires: clues from psychological type. *Mental Health, Religion and Culture*, *18*(7), 566–575. https://doi.org/10.1080/13674676.2014.961248.

Rao, K., & Pennington, J. (2013). Should the third reminder be sent? The role of survey response timing on web survey results. *International Journal of Market Research*, *55*(5), 651–674. https://doi.org/10.2501/ijmr-2013-056.

Rashidian, A., van der Meulen, J., & Russell, I. (2008). Differences in the contents of two randomized surveys of GPs' prescribing intentions affected response rates. *Journal of Clinical Epidemiology*, *61*(7), 718–721. https://doi.org/10.1016/j.jclinepi.2007.10.019.

Reips, U.-D. (2013). Virtual experiments: a psychological laboratory on the Internet. In C. Hine (Ed.), *Virtual research methods*. London: SAGE. ISBN 978-0-85702-740-5.

Reips, U.-D. (2021). Web-based research in psychology. *Zeitschrift für Psychlogie*, *229*(4), 198–213. https://doi.org/10.1016/j.newideapsych.2007.10.001.

Reips, U.-D., & Franek, L. (2004). Mitarbeiterbefragung per Internet oder Papier? Der Einfluss von Anonymität, Freiwilligkeit und Alter auf das Antwortverhalten [Employee survey via Internet or paper? The influence of anonymity, voluntariness and age on response behavior]. *Wirtschaftspsychologie*, *1*, 67–83.

Rueegg, C.S., Gianinazzi, M.E., Michel, G., Zwahlen, M., Von Der Weid, N.X., Kuehni, C.E., Ammann, R., Angst, R., Ansari, M., Beck Popovic, M., Brazzola, P., Greiner, J., Grotzer, M., Hengartner, H., Kuehne, T., Leibundgut, K., Niggli, F., & Rischewski, J. (2017). No evidence of response bias in a population-based childhood cancer survivor questionnaire survey—results from the Swiss childhood Cancer Survivor study. *PLoS ONE*, *12*(5), e176442. https://doi.org/10.1371/journal.pone.0176442.

Schmidt, F.L. & Hunter, J.E. (2014). *Methods of meta-analysis: correcting error and bias in research findings* (3rd edn.). SAGE. https://doi.org/10.4135/9781483398105.

Shih, T.-H., & Fan, X. (2007). Response rates and mode preferences in web-mail mixed-mode surveys: a meta-analysis. *International Journal of Internet Science*, *2*(1), 59–82.

Siddaway, A.P., Wood, A.M., & Hedges, L.V. (2019). How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, *70*, 747–770. https://doi.org/10.1146/annurev-psych-010418-102803.

Sigman, R., Lewis, T., Yount, N.D., & Lee, K. (2014). Does the length of fielding period matter? Examining response scores of early versus late responders. *Journal of Official Statistics*, *30*(4), 651–674. https://doi.org/10.2478/JOS-2014-0042.

Slavin, R.E. (1986). Best-evidence synthesis: an alternative to meta-analytic and traditional reviews. *Educational Researcher*, *15*(9), 5–11. https://doi.org/10.3102/0013189X015009005.

Smyth, J.D., Dillman, D.A., Christian, L.M., & McBride, M. (2009). Open-ended questions in web surveys: can increasing the size of answer boxes and providing extra verbal instructions improve response quality? *Public Opinion Quarterly*, *73*(2), 325–337. https://doi.org/10.1093/poq/nfp029.

Tourangeau, R., Rips, L., & Rasinski, K. (2000). Mode of data collection. In *The psychology of survey response* (pp. 289–312). Cambridge University Press. https://doi.org/10.1017/CBO9780511819322.011.

Venette, S., Sellnow, D., & McIntyre, K. (2010). Charting new territory: assessing the online frontier of student ratings of instruction. *Assessment & Evaluation in Higher Education*, *35*(1), 101–115. https://doi.org/10.1080/02602930802618336.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Viechtbauer, W., & Cheung, M.W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125.

Vink, J.M., & Boomsma, D.I. (2008). A comparison of early and late respondents in a twin-family survey study. *Twin Research and Human Genetics*, *11*(2), 165–173. https://doi.org/10.1375/twin.11.2.165.

Voigt, L.F., Koepsell, T.D., & Daling, J.R. (2003). Characteristics of telephone survey respondents according to willingness to participate. *American Journal of Epidemiology*, *157*(1), 66–73. https://doi.org/10.1093/aje/kwf185.

Witry, M.J., Arya, V., Bakken, B.K., Gaither, C.A., Kreling, D.H., Mott, D.A., Schommer, J.C., & Doucette, W.R. (2021). National Pharmacist Workforce Study (NPWS): description of 2019 survey methods and assessment of nonresponse bias. *Pharmacy*, *9*(1), 20. https://doi.org/10.3390/pharmacy9010020.

Yan, T., Tourangeau, R., & Arens, Z. (2004). When less is more: are reluctant respondents poor reporters? In *Survey research methods section, annual meetings of the American Statistical Association* (pp. 4632–4651). http://www.amstat.org/Sections/Srms/Proceedings/y2004/files/Jsm2004-000169.pdf%0A0768605962tsPoorReporters.pdf.

Zachry, W.M., Dalen, J.E., & Jackson, T.R. (2003). Clinicians' responses to direct-to-consumer advertising of prescription medications. *Archives of Internal Medicine*, *163*(15), 1808–1812. https://doi.org/10.1001/archinte.163.15.1808.

Appendix

Table A1

Summary of the Data Quality Differences between Early and Late Respondents

Study	Mode 1 = mail 2 = web	Late response definition	Data quality indicator (0 = no difference, 1 = difference, 2 = mixed results)						Comments
			Item non-response	Breakoff	Quality of open-ended questions	Motivated Misreporting	Non-substantive answers	Equivalence	
Brehaut et al. (2006)	1	Extreme group	1	–	–	–	–	–	Significant effect for those who responded after the third reminder compared to those who responded earlier
Friedman et al. (2003)	1	Threshold	2	–	–	–	2	–	Significant effects for several variables for 2 out of 3 survey waves
Helasoja et al. (2002) (1)	1	Threshold	2	–	–	–	–	–	No significant effects, but consistently higher item nonresponse for late respondent for all four(*) countries
Helasoja et al. (2002) (2)	1	Threshold	2	–	–	–	–	–	
Helasoja et al. (2002) (3)	1	Threshold	2	–	–	–	–	–	
Hox et al. (2012)	1	Threshold	–	–	–	–	–	0	No differences in measurement structure
Korkeila et al. (2001)	1	Threshold	2	–	–	–	–	–	Only significant differences for the request for consent as well as for the sensitive topics were found
Bach et al. (2020) (1)	2	Extreme group	–	–	–	2	–	–	There seems to be small evidence for a connection between respondents' reluctance and motivated misreporting
Bach et al. (2020) (2)	2	Extreme group	–	–	–	0	–	–	

Table A1

(Continued)

Study	Mode 1 = mail 2 = web	Late response definition	Data quality indicator (0 = no difference, 1 = difference, 2 = mixed results)						Comments
Göritz and Stieger (2009) (1)	2	Continous	0	2	–	–	–	–	In Exp. 1, the sooner people responded if they had been given a tight deadline the more likely they completed the survey
Göritz and Stieger (2009) (2)	2	Continous	0	0		–			
Göritz and Stieger (2009) (3)	2	Continous	0	0	–	–	–	–	
Göritz and Stieger (2009) (4)	2	Continous	0	0		–			
Rao and Pennington (2013)	2	Extreme group	1	1	–	–	–	–	The differences in missing data and the break-off rate among early, intermediate, and late respondents were significant
Smyth et al. (2009)	2	Threshold	–	–	1	–	–	–	Only late respondents are influenced by box size, with those receiving the small box giving significantly shorter responses than those receiving the long box
Smyth et al. (2009)	2	Threshold	–	–	0	–	–	–	No differential effects between both groups were found for motivating instructions
Smyth et al. (2009)	2	Threshold	0	–	2	–	–	–	The "important" introduction compared to no introduction significantly increased both themes and elaboration among late but not early respondents

Note: (*) In Estonia, the information of response round was not available. Therefore, these study results were not included in the systematic review and meta-analysis.

Table A2

Meta-regression of Potential Moderator Variables on the Target Population/Sample and the Survey Design (Including Dichotomized Variables)

	Web					Mail					Total sample				
Variable	N	B	Se	p	R2	N	B	Se	p	R2	N	B	Se	p	R2
Target Population & Sample															
Experts: yes	17	0.09	0.06	0.12	0.00	47	0.05	0.07	0.43	0.69	64	0.06	0.05	0.20	0.00
Sample: General Population	17	–0.12	0.03	0.00[a]	70.40	47	–0.06	0.06	0.31	2.90	64	–0.07	0.03	0.02[a]	38.80
Response rate: > 50%	17	0.06	0.03	0.06	54.20	44	0.00	0.06	0.97	0.46	61	0.02	0.04	0.61	11.30
Country: North America	17	0.16	0.04	0.00[a]	19.20	47	–0.15	0.05	0.007[a]	31.46	64	–0.04	0.04	0.30	1.83
Field: Health	17	–0.10	0.05	0.02[a]	7.60	47	0.00	0.06	0.92	16.60	64	–0.04	0.04	0.35	3.20
Field: Method	17	0.15	0.03	<0.0001[a]	64.90	47	–0.04	0.07	0.59	27.60	64	0.04	0.05	0.37	9.00
Survey design															
Field period: 57 and more days	15	0.18	0.05	0.002[a]	1.56	41	–0.04	0.04	0.33	9.80	56	0.00	0.04	0.93	1.60
Interval to first reminder: 0 to 14 days	–1	–	–	–	–	39	–0.03	0.04	0.45	10.60	51	–0.01	0.04	0.68	10.87
Survey lenght: medium and long	13	–0.05	0.03	0.16	63.40	25	0.06	0.05	0.22	26.10	38	0.00	0.03	0.88	48.57
Reminders: 3 and more	14	0.03	0.07	0.50	64.90	47	–0.05	0.06	0.44	0.54	61	–0.02	0.05	0.59	1.26
Personalized: yes	17	0.02	0.05	0.62	5.00	47	–0.02	0.06	0.73	25.09	64	0.00	0.04	0.98	0.85
Incentive: yes	17	0.05	0.05	0.31	6.80	47	–0.03	0.05	0.66	24.10	64	0.00	0.04	0.95	0.60
Mode											64	–0.02	0.04	0.61	22.40

1: Only data for one of the two values are available.
Note: [a] Analyses revealing a significant impact.

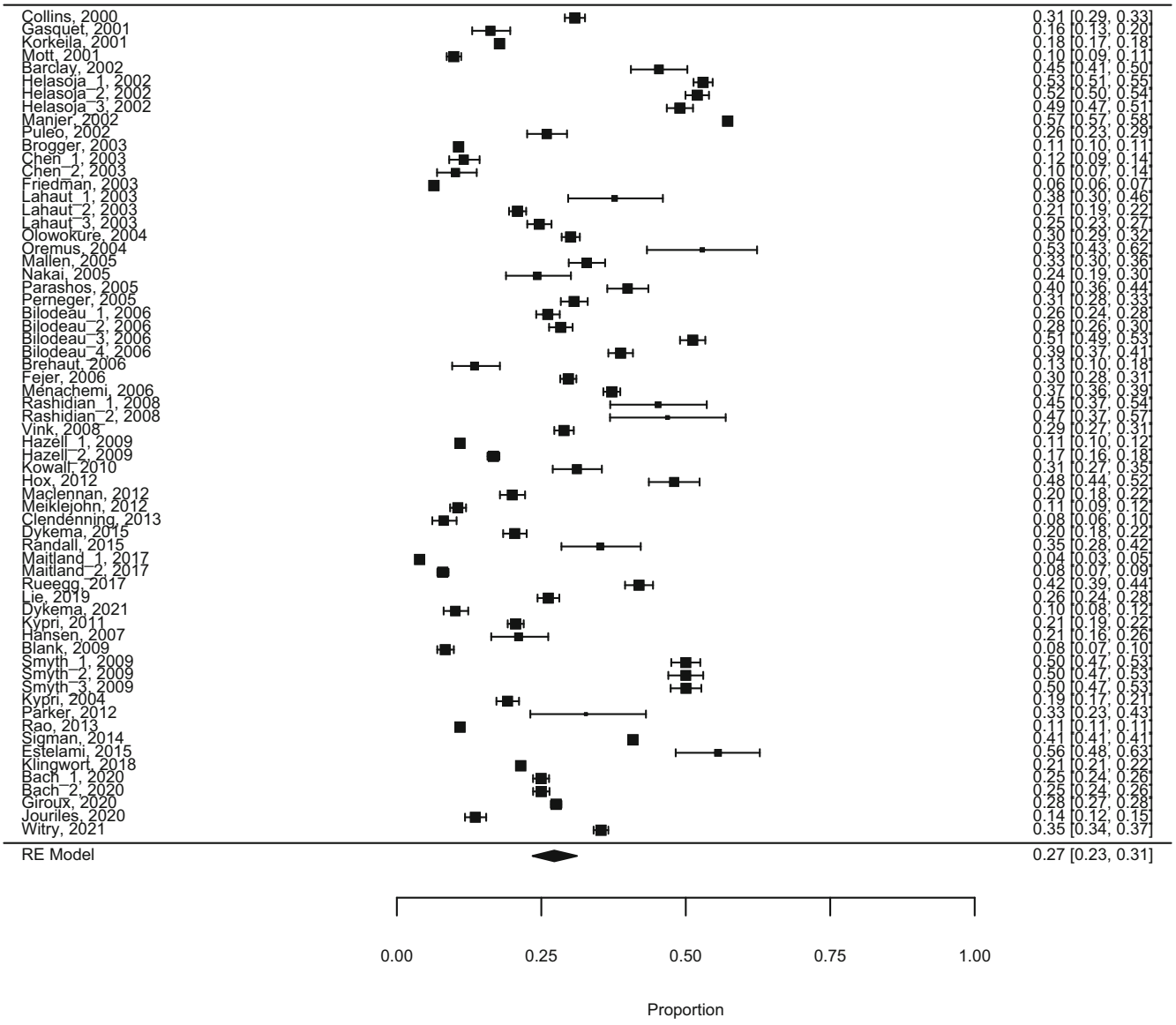

Fig. A1

Forest Plot of All Considered Studies (n = 64) Note: This forest plot includes all considered studies (n = 64). On the left side, you see the first authors' name of each study and the publication year. On the right side you see for each study the portion of late respondents and the associated confidence interval. In the middle, you see each study with their confidence interval plotted. At the bottom, the aggregated meta-analytic value across the 64 studies is plotted represented by an average portion of 27% of late respondents across studies

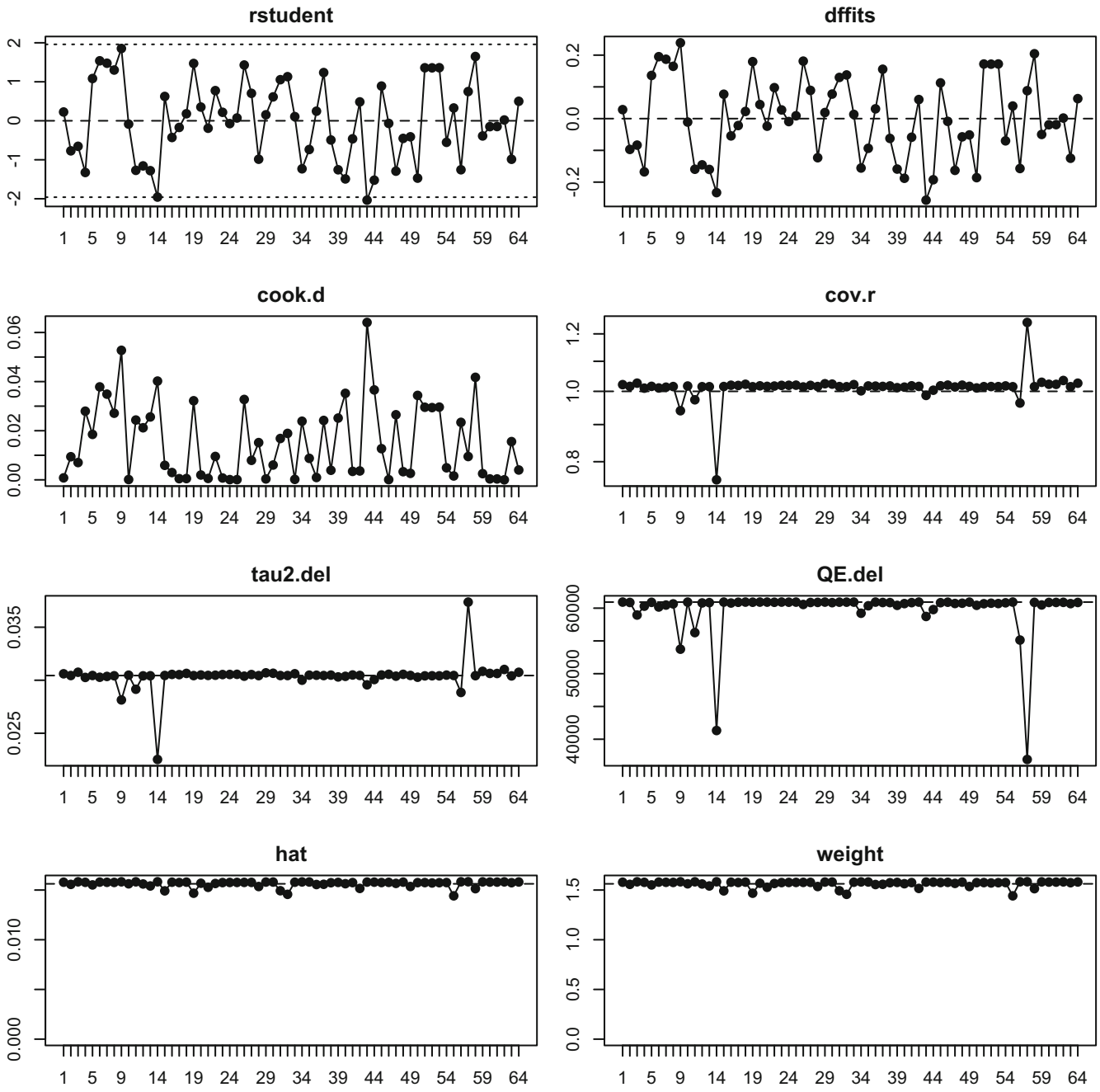

Fig. A2

Different Outlier Analyses on the Overall Database of 64 Studies (N) Note: According to Viechtbauer and Cheung (2010) we presented a plot of the externally standardized residuals, DFFITS values, Cook's distances, covariance ratios, estimates of τ 2 and test statistics for (residual) heterogeneity. These different outlier analyses considering the complete dataset of 64 studies were conducted. As you see from the different plots, no study is marked by a red dot, hence, there is no outlier study revealed. For further information about outlier analysis, we refer the interested reader to the well described overview in Viechtbauer and Cheung (2010)

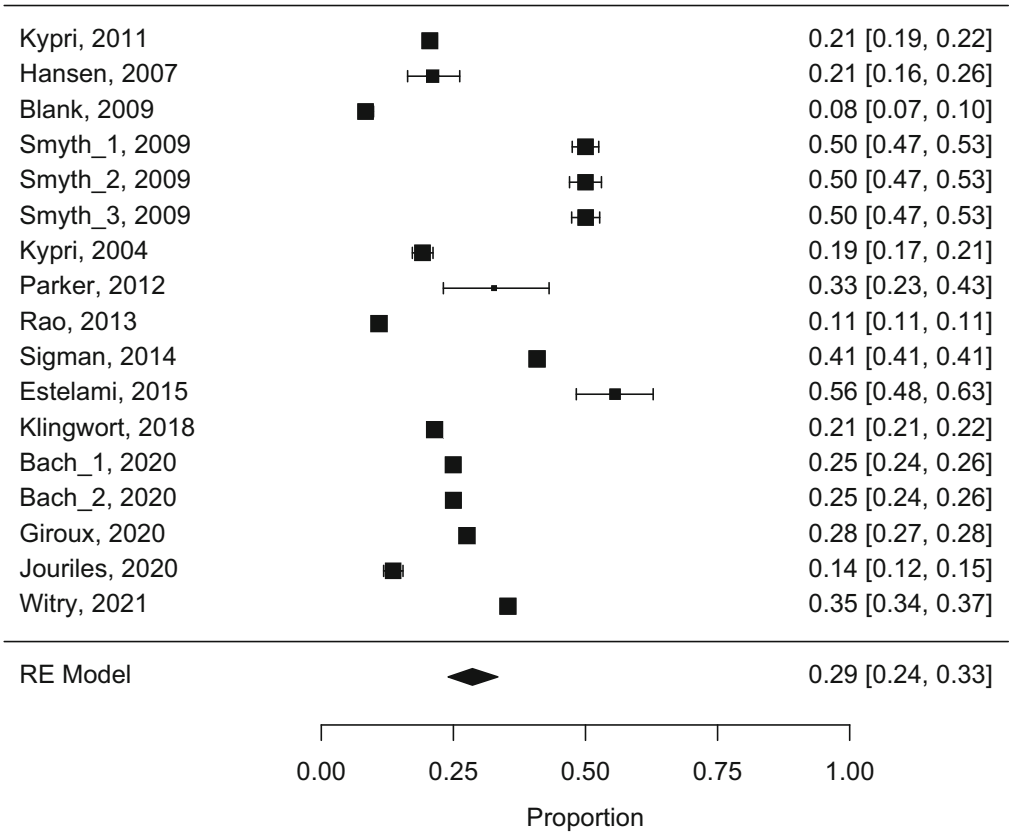

Fig. A3

Forest Plot of the Subsample of Studies with Web Surveys (n = 17) Note: This forest plot includes all web studies (n = 17). On the left side, you see the first authors' name of each study and the publication year. On the right side you see for each study the portion of late respondents and the associated confidence interval. In the middle, you see each study with their confidence interval plotted. At the bottom, the aggregated meta-analytic value across the 17 studies is plotted represented by an average portion of 29% of late respondents across studies

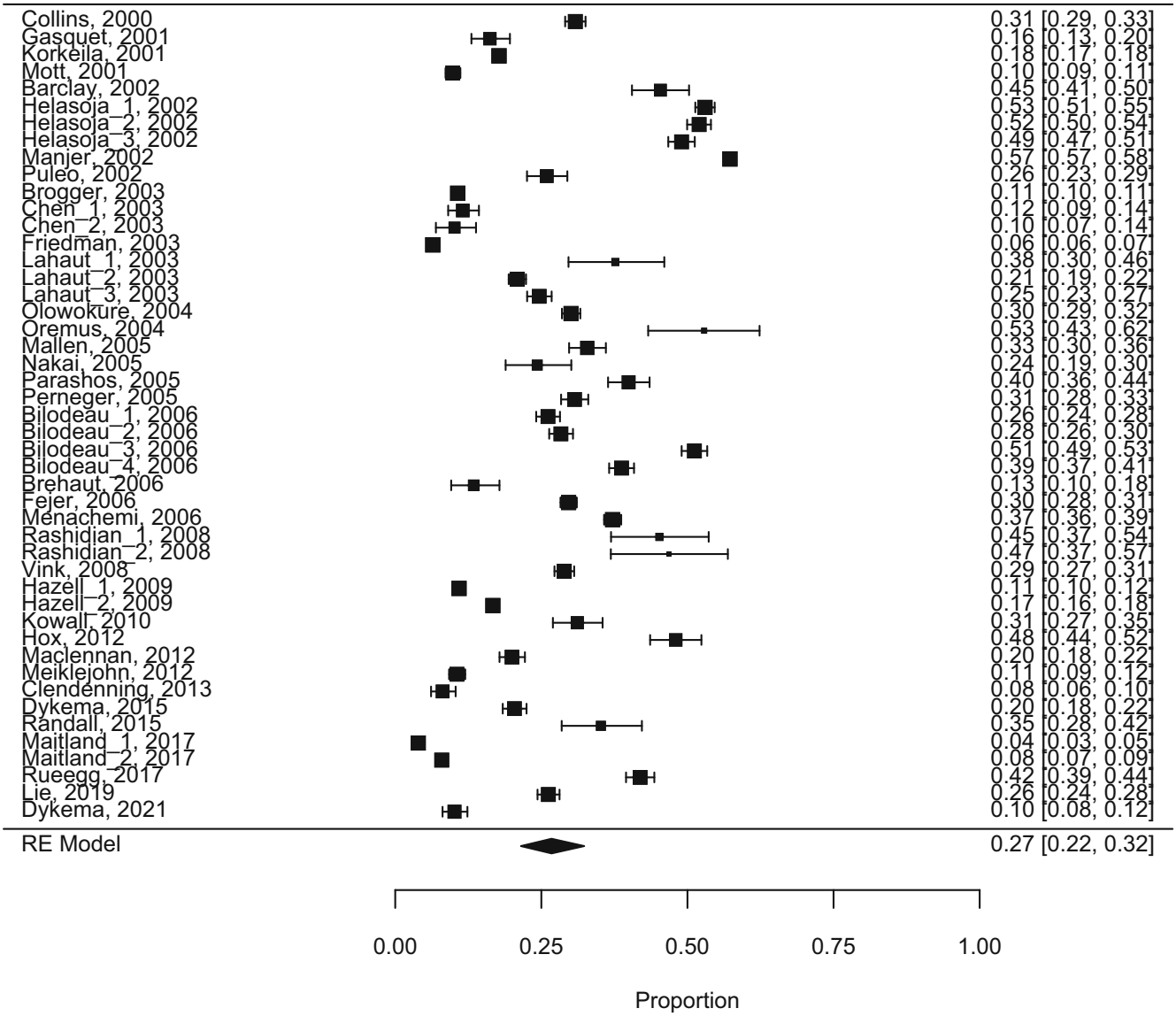

Fig. A4

Forest Plot of the Subsample of Studies with Mail Surveys (n = 47) Note: This forest plot includes all web studies (n = 47). On the left side, you see the first authors' name of each study and the publication year. On the right side you see for each study the portion of late respondents and the associated confidence interval. In the middle, you see each study with their confidence interval plotted. At the bottom, the aggregated meta-analytic value across the 47 studies is plotted represented by an average portion of 27% of late respondents across studies