Acta Informatica Pragensia
2026, Volume 15, Issue 1, 173–197
https://doi.org/10.18267/j.aip.279

PRAGUE UNIVERSITY
OF ECONOMICS
AND BUSINESS

Review Open Access

Data Quality in Estimates from Probability-Based Online Panels: Systematic Review and Meta-Analysis

Andrea Ivanovska [1], **Michael Bosnjak** [1,2], **Vasja Vehovar** [1]

[1] Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia
[2] Department of Psychological Research Methods, Trier University, Trier, Germany

Corresponding author: Andrea Ivanovska (ai67517@student.uni-lj.si)

Editorial Record

First submission received:
June 4, 2025

Revision received:
August 14, 2025

Accepted for publication:
August 31, 2025

Academic Editor:
Stanislava Mildeova
University of Finance and Administration, Czech Republic

This article was accepted for publication by the Academic Editor upon evaluation of the reviewers' comments.

How to cite this article:
Ivanovska, A., Bosnjak, M., & Vehovar, V. (2026). Data Quality in Estimates from Probability-Based Online Panels: Systematic Review and Meta-Analysis. *Acta Informatica Pragensia*, 15(1), 173–197. https://doi.org/10.18267/j.aip.279

Abstract

Background: General population surveys now increasingly use nonprobability samples from access panels instead of probability-based methods, which often leads to lower-quality estimates. In response, many official and academic surveys have adopted probability-based online panels (PBOPs), which use probability sampling and retain participants for follow-up surveys. While these panels reduce costs compared to one-time surveys, they still face low response rates and other challenges that may affect data quality.

Objective: This study aimed to assess the accuracy of PBOPs by synthesising evidence on relative bias (RB), and to examine how RB varies by country, domain, measurement level, and item sensitivity.

Methods: A systematic review yielded 44 eligible studies from 12 countries, and 1,897 effect sizes of absolute RB from studies that compared PBOP estimates to benchmarks. A three-level random effects meta-analytic model accounted for variance across studies, within studies and sampling variance. Moderator analyses evaluated the influence of country, item topic, measurement level and sensitivity on RB. Sensitivity analyses excluded the top 5% of RB outliers to test robustness.

Results: The pooled RB was 23.14% (95% CI: 18.38%–27.91%) and heterogeneous. Most variance was attributed to within-study item-level differences. Country and topic did not significantly moderate RB. Items with high topic sensitivity had significantly higher RB (+19.33%) than items with no sensitivity. Ordinal items had significantly lower RB than nominal (–14.90%). However, when sensitivity and measurement level were modelled together, substantial residual heterogeneity remained.

Conclusion: While PBOPs offer cost and logistical advantages, they require careful design considerations to lower substantial bias, especially regarding item sensitivity and measurement scale. PBOPs may not be suitable for certain question types, like sensitive or low-prevalence behaviours, especially when high accuracy is needed. Improved methodological planning and innovations are needed to improve PBOP data quality.

Index Terms

Online surveys; Probability-based online panels; Data quality; Relative bias; Meta-analysis.

1 INTRODUCTION

The interaction between information and communication technologies (ICTs) and society drives transformative changes at individual, organisational, and societal levels (Rebenok et al., 2024; Vehovar et al., 2022). These developments began in the 1970s and accelerated with the rise of the internet, smart devices, and artificial intelligence, driven largely by the aim to reduce expensive and error-prone human labour. This trend is evident across various sectors, from robotics to automated services in banking, retail and public administration, where interpersonal interactions have been replaced by automated procedures (Dillman, 2007).

Within this context, social science research methodology—particularly survey data collection, the focus of this article—has moved from computer-assisted telephone interviewing in the 1970s to computer-assisted personal interviewing in the 1980s and online surveys in the 1990s (de Leeuw & Nicholls, 1996; Symbaluk & Hall., 2024), which was driven by efficiency, flexibility, and data quality (Berzelak & Vehovar, 2009; Dillman, 2007). This transition, reflecting technological progress and evolving research demands (Vehovar & Lozar Manfreda, 2017), has made online surveys the dominant mode due to their speed, flexibility, and cost-effectiveness. Nevertheless, the science of survey methodology was originally established in the environment of traditional methods such as face-to-face interviews, telephone surveys, and mail surveys (de Leeuw, 2008) central to large-scale, institutionalized data collection efforts, which relied on systematic measurement, standardized procedures and professional staff to ensure credibility and national impact (Groves et al., 2009). These methodologies required high response rates, broad population coverage, and probability sampling, where each unit had a known, nonzero chance of selection (Stadtmüller et al., 2023). General population surveys are often based on population registers or address-based sampling frames, aiming for near-full coverage and response rates of 70–80% (Callegaro et al., 2015c; Groves et al., 2009), considered the gold standard. Over time, participation has declined due to rising operational costs (interviewer wages, travel, logistics), increased recruitment effort (Ormston et al., 2024), respondent fatigue, reduced landline use, and heightened privacy concerns (Boland et al., 2006; Vehovar & Beullens, 2017).

The shift towards peopleless and paperless survey data collection represents a major transformation in survey research (Dillman, 2007; Tourangeau et al., 2013). While replacing human interviewers and paper instruments with digital technologies has improved efficiency, it has not solved the problem of declining participation; in fact, additional challenges have emerged (Vehovar et al., 2002; e.g., Ryan et al., 2024). Three decades of online survey developments have highlighted persistent issues with noncoverage and nonresponse (Berzelak et al., 2025). For example, Gaia et al. (2025) find that, despite rising Internet use , a non-negligible share of the population remains offline and coverage bias varies across countries (Gaia et al., 2025). In digital environments, individuals may experience especially low willingness to participate, perceived loss of control and information overload (van der Schyff et al., 2023). Furthermore, it has become increasingly difficult to build the trust needed to establish participant cooperation (Pang & Capek, 2020). Operating within centralised, privacy-fatigued digital spaces, online surveys require decentralised, trust-based designs (Vojíř & Kučera, 2021). Yet, they still face substantially lower response rates (Daikeler et al., 2020; Lozar Manfreda et al., 2008), sometimes in single digits, making probability-based online surveys resemble nonprobability ones (Vehovar et al., 2016).

Transitioning to online survey methods thus requires more than simply digitalising paper questionnaires, as replicating the social, emotional and psychological dynamics of in-person methods remains a significant challenge (Pang & Capek, 2020). One reaction to these challenges has been the development of panels, in which participants are recruited once and give consent for ongoing participation, allowing them to be re-contacted for future surveys in exchange for incentives (Callegaro et al., 2015b). Particularly in digital contexts, surveys should not function solely as a one-off data collection tool, but as ongoing, trust-based engagements embedded within broader data infrastructures (Lorenz & Konečný, 2023). Reflecting this shift, online panels have gained popularity as efficient data collection tools that streamline fieldwork, reduce costs and enable timely data delivery (Callegaro et al., 2015b).

However, due to cost pressures, most online panels are nonprobability online panels, which rely on nonprobability recruitment methods, such as advertising and convenience sampling (Baker et al., 2010), recruiting a group of respondents to participate in surveys continuously. These access, opt-in, or volunteer panels dominate survey research in marketing and business contexts and are increasingly used in academic and government studies. While cost-effective and fast (Räsänen et al., 2024), they generally yield lower data quality than probability samples, as valid statistical inference requires a known probability of selection (Lavrakas et al., 2022). In response, academic and official institutions have developed probability-based online panels (PBOPs; see Appendix A for glossary-style definitions for key terms), which use traditional probability sampling to improve representativeness (Blom et al., 2016; Cornesse et al., 2022a). Initial recruitment methods may include single-mode (e.g., postal or face-to-face) or mixed-mode approaches (e.g., mail and face-to-face), while data collection can be conducted exclusively online or via mixed modes (e.g., web in combination with mail or face-to-face) (Bosch & Maslovskaya, 2023). In this study, we define PBOPs as any panel using probability-based sampling with primarily online data collection. Compared to stand-alone online probability surveys, which require own and separate recruitment, PBOPs reduce costs by avoiding repeated recruitment. Nevertheless, they remain more expensive than nonprobability panels, because the available probability sampling frames require traditional recruitment modes. As costs are closely tied to data quality—particularly to the response rates—researchers must carefully balance expenditure and data quality to meet their specific needs. While estimating costs is relatively straightforward (e.g., by consulting contractors or conducting internal calculations), assessing data quality is more complex.

Within this context, the present study examines response data quality in PBOPs. In recent years, the establishment of several new PBOPs has led to renewed interest in their methodology and to deeper questioning of their role within the wider survey landscape (Bosnjak et al., 2016). Despite this, their evaluations are either fragmented across specific panels or limited in scope. For example, Maslovskaya and Lugtig (2022) assessed representativeness over time and across countries, but only within a single cross-national PBOP (CRONOS). Blom et al. (2016) qualitatively compared four European PBOPs, detailing recruitment strategies, retention practices, and panel structures, but did not assess estimate accuracy or bias. More comprehensive reviews offer only partial insights. Kocar & Kaczmirek (2023) conducted a meta-analysis of recruitment outcomes from 23 PBOPs worldwide, reporting an average overall recruitment rate below 20%, but focused solely on recruitment rather than quality of the resulting estimates. Bosch and Maslovskaya (2023) conducted a comprehensive literature review using the Total Survey Error framework to compare PBOPs with other survey modes and among themselves, but did not assess or quantify the bias of the estimates. Therefore, a clear gap remains: no study has systematically reviewed the published evidence on estimate bias from different PBOPs across. This study addresses that gap by conducting a systematic literature review, followed by a corresponding meta-analysis. To maintain a manageable scope, the review focuses specifically on bias of the estimates, which is one of the key indicators of data quality in survey research.

1.1 Background

1.1.1 Measuring data quality of PBOPs

Numerous elements and metrics are available for evaluating PBOPs, including recruitment rate, profile rate, completion rate, cumulative response rate, attrition rate, design effect and R indicators (DiSogra & Callegaro, 2015). PBOP data quality is typically evaluated by comparing PBOP data with an external benchmark, typically a government or other high-quality traditional survey (e.g., Cornesse & Blom, 2023). In this context, a benchmark is a trusted reference value, usually from a high-quality source, against which PBOP estimates are compared to assess accuracy. Benchmark comparisons help identify potential biases in PBOPs, though mode effects and benchmark comparability require careful consideration. Common accuracy metrics for data quality include direct comparisons of response distributions to benchmarks, reporting the lowest and highest values across panels, computing the average estimate across panels and assessing error metrics such as average absolute error, largest absolute error and the number of significant differences from a benchmark (Callegaro et al., 2014). These metrics can be reported as weighted or unweighted and are used to determine the extent of bias in panel estimates compared to high-quality benchmark surveys.

Studies evaluating PBOPs commonly report absolute percentage differences between survey estimates and external benchmarks (e.g., Kocar & Baffour, 2023; Mercer & Lau, 2023). However, this approach can be misleading, as it does not account for variation in scale across estimates, complicating comparisons (Eckman, 2015). Instead, the absolute value of relative bias (RB) is a metric that expresses the absolute percentage difference between a survey estimate and its benchmark as a proportion of the benchmark value, addressing this limitation (Eckman et al., 2023). This metric facilitates more meaningful comparisons across variables and populations (Eckman, 2015).

1.1.2 Research questions and aims

A meta-analytic approach is well-suited for evaluating the data quality of PBOPs because it enables the quantitative synthesis of results across multiple studies to produce a single, interpretable estimate of effect (Harrer et al., 2021). In contrast to narrative reviews, which may be influenced by subjective interpretation, meta-analyses apply transparent rules for study selection, data extraction and synthesis, which increases the reliability of conclusions. So, building on the preceding context, this study addresses the following research questions within an exploratory meta-analytic framework:

Research Question 1: What is the extent of RB in estimates from PBOP surveys compared to benchmark surveys?

The first objective is to estimate and summarise the average RB across PBOP-based survey estimates. We also assess whether these effects are homogeneous. However, given the numerous survey design and contextual factors that influence data quality, we anticipate considerable heterogeneity.

Research Question 2: To what extent do factors such as domain, measurement level, sensitivity and country-level differences moderate RB in PBOP estimates?

To explore this question, we examine the following moderators:

> *Domain*. Data quality may vary by topic, as topics differ in familiarity, sensitivity and social desirability. Topic-related characteristics—such as domain and conceptual focus—are key predictors of both reliability and validity (Felderer et al., 2024). Certain topics are likely to elicit more consistent and accurate responses, and we, therefore, expect substantial variation in RB by domain.
>
> *Measurement level*. The level of measurement can affect data quality. Nominal variables classify responses into categories without order; ordinal variables introduce ranking but with unequal intervals, which may lead to interpretive ambiguity (Lalla, 2017). We hypothesise that nominal and ordinal variables, lacking the precision of interval measures, may be more susceptible to subjective interpretation and bias.
>
> *Country*. Cultural norms shape survey responses: collectivist cultures may heighten social desirability bias in public behaviour reporting but improve recall, while individualist cultures may encourage independent responding and greater accuracy for private behaviours (Schwarz et al., 2008; Ji et al., 2000). Differences in internet coverage, digital literacy, and access (Hernandez & Faith, 2023) may further affect response rates and comparability, so we expect PBOP response patterns to vary across countries.
>
> *Sensitivity*. Sensitivity affects data quality as respondents may underreport undesirable behaviours or overreport desirable ones. While online surveys can reduce social desirability bias by increasing perceived privacy (Bosch & Maslovskaya, 2023), they may also increase item nonresponse for sensitive questions (Goodman et al., 2022) and reduce engagement for complex items. We therefore expect sensitive items to show greater bias.

2 METHODOLOGY

2.1 Literature selection

This systematic review and meta-analysis followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Page et al., 2021) guidelines to systematically identify, screen and select relevant studies. The search was conducted using the Digital Library of the University of Ljubljana (DiKUL), which indexes 155 databases and information sources, including Web of Science, Scopus and PubMed (Centralna tehniška knjižnica Univerze v Ljubljani, 2025). The full search process is illustrated in Figure 1.

The search strategy targeted studies examining PBOPs and their data quality. To construct a comprehensive search string, two groups of terms were used: one related to PBOPs and the other to data quality. PBOP terms included “probability panel,” “probability-based panel,” “probability online panel,” “probability-based online panel,” “probability web panel,” “probability-based web panel,” “probability internet panel,” and “probability-based internet panel.” Data quality terms included “difference,” “evaluation,” “comparison,” “data quality,” “bias,” “error,” and “accuracy.” Within each group, terms were combined using ‘OR’ and the two groups were joined using ‘AND’ to identify studies referencing both a PBOP and a data quality or bias-related measure. A citation analysis was also conducted by reviewing the reference lists of eligible studies to identify additional relevant research.

Studies were included if they:

1. Were related to PBOPs.
2. Compared PBOP estimates with external benchmarks. For this review, a valid external benchmark was defined as official statistics or any other survey relying on traditional survey methods.
3. Provided RB measures or included data enabling its calculation.

Studies were excluded if they:

1. Did not focus on PBOPs.
2. Did not compare PBOP estimates with external benchmarks.
3. Did not include empirical data required to calculate RB.

When multiple benchmarks were available for the same estimate, the first reported benchmark was used for analysis.

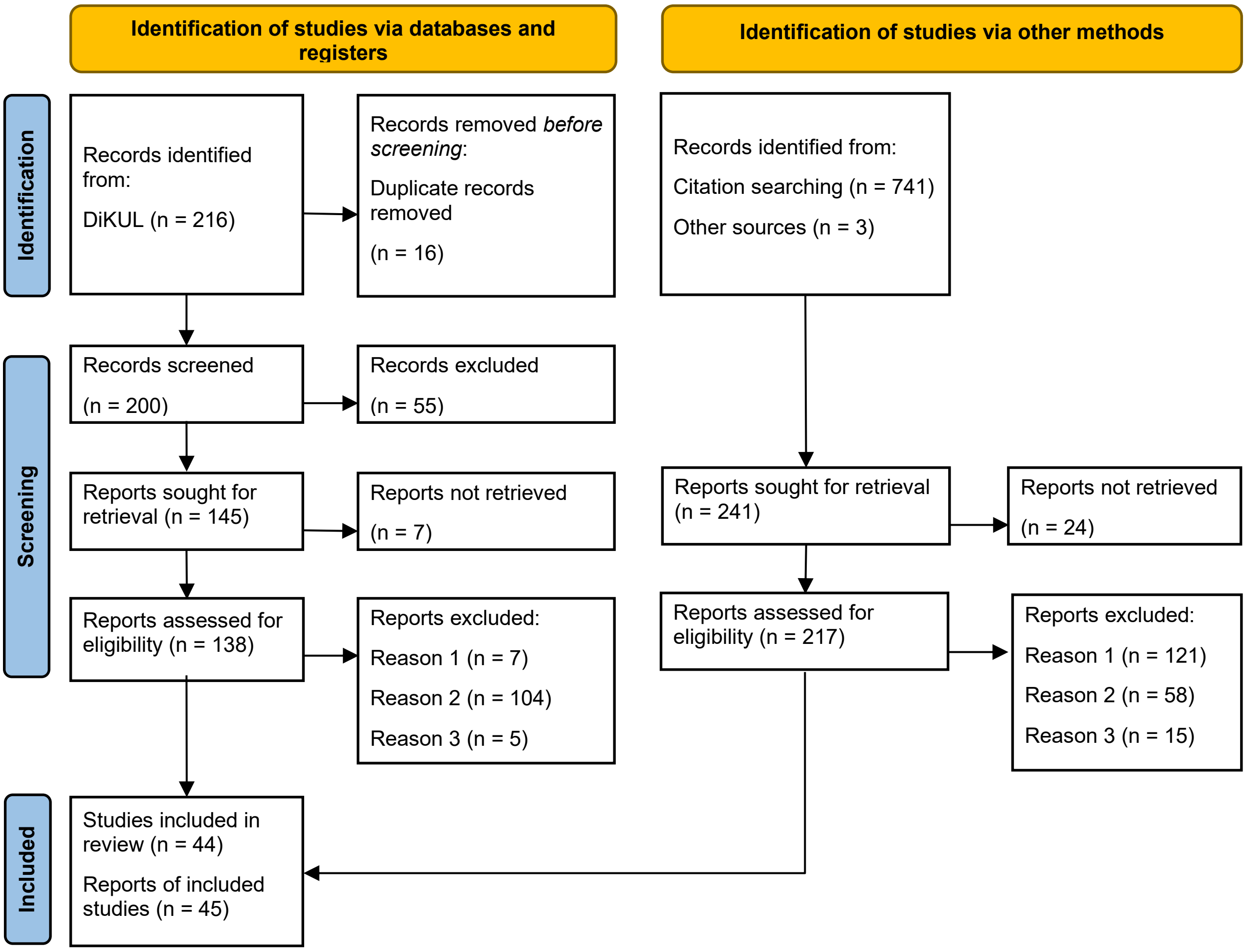
Identification of studies via databases and registers
Identification of studies via other methods
Identification
Screening
Included
Records identified from:
DiKUL (n = 216)
Records removed *before screening*:
Duplicate records removed
(n = 16)
Records identified from:
Citation searching (n = 741)
Other sources (n = 3)
Records screened
(n = 200)
Records excluded
(n = 55)
Reports sought for retrieval (n = 145)
Reports not retrieved
(n = 7)
Reports sought for retrieval (n = 241)
Reports not retrieved
(n = 24)
Reports assessed for eligibility (n = 138)
Reports excluded:
Reason 1 (n = 7)
Reason 2 (n = 104)
Reason 3 (n = 5)
Reports assessed for eligibility (n = 217)
Reports excluded:
Reason 1 (n = 121)
Reason 2 (n = 58)
Reason 3 (n = 15)
Studies included in review (n = 44)
Reports of included studies (n = 45)

Figure 1. *Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram showing the literature selection process.*

A total of 44 studies met the inclusion criteria for this systematic review and meta-analysis (see Table 1). The United States was the most frequently represented country, accounting for over half of all studies. Within the U.S., Knowledge Networks or Knowledge Panel was the most commonly examined panel, appearing in at least 11 studies. Other U.S.-based studies included AmeriSpeak (Bilgen et al., 2018), Axios-Ipsos (Bradley et al., 2021), the RAND American Life Panel (Schonlau et al., 2007) and the TCS Panel (Liu et al., 2022). Three studies did not disclose the panel name (Mercer & Lau, 2023; Unangst et al., 2020; Yeager et al., 2011), reflecting reduced transparency in reporting.

Table 1. *Description of included studies.*

Study	**Panel**	**Country**
Arcos et al. (2020)	PACIS	Spain
Bell et al. (2011)	Knowledge Networks	USA
Berrens et al. (2003)	Knowledge Networks	USA
Bilgen et al. (2018)	AmeriSpeak	USA

Study	Panel	Country
Blom et al. (2015)	German Internet Panel	Germany
Blom et al. (2017)	German Internet Panel	Germany
Bottoni and Fitzgerald (2021)	CRONOS	Estonia, Great Britain and Slovenia
Bradley et al. (2021)	Axios-Ipsos	USA
Chang and Krosnick (2009)	Knowledge Networks	USA
Cho et al. (2017)	KAMOS	South Korea
Cornesse & Schaurer (2021)	German Internet Panel	Germany
	GESIS Panel	Germany
Cornesse et al. (2022b)	German Internet Panel	Germany
	Mannheim Corona Study	Germany
Dever et al. (2021)	KnowledgePanel	USA
Dickie et al. (2007)	Knowledge Networks	USA
Grönlund & Strandberg (2014)	eOpinion	Finland
Hemsworth et al. (2021)	MyView	Australia
Herman et al. (2024)	KnowledgePanel	USA
Høgestøl & Skjervheim (2014)	Norwegian citizen panel	Norway
Huggins et al. (2001)	Knowledge Networks	USA
Kaczmirek et al. (2019)	Life in Australia	Australia
Kaufman et al. (2016)	KnowledgePanel	USA
Kennedy et al. (2016)	American Trends Panel	USA
Kocar & Biddle (2023)	Life in Australia	Australia
Lee (2006)	Knowledge Networks	USA
Leenheer & Scherpenzeel (2013)	LISS Panel	The Netherlands
Liu et al. (2022)	TCS Panel	USA
Lugtig et al. (2014)	LISS Panel	The Netherlands
MacInnis et al. (2018)	Knowledge Networks	USA
McMillen et al. (2013)	Knowledge Networks	USA
Mercer & Lau (2023)	Unnamed PBOP	USA
	Unnamed PBOP	USA
	Unnamed PBOP	USA
Pennay et al. (2018)	ANU Poll	Australia
Revilla (2013)	LISS Panel	The Netherlands
Scherpenzeel and Bethlehem (2011)	LISS Panel	The Netherlands
Schonlau et al. (2007)	RAND American Life Panel	USA
Seol et al. (2023)	Gallup Korea's Online Panel	South Korea
Smith (2003)	Knowledge Networks	USA
Smith et al. (2004)	Knowledge Networks	USA
Spijkerman et al. (2009)	Dutch online panel of Survey Sampling International LLC	The Netherlands
Stanley et al. (2020)	Knowledge Panel	USA
Struminskaya et al. (2014)	GESIS Online Panel Pilot	Germany
Struminskaya et al. (2016)	GESIS Online Panel Pilot	Germany
Unangst et al. (2020)	Unnamed PBOP	USA
	Unnamed PBOP	USA
Vaithianathan et al. (2021)	The Singapore Life Panel	Singapore
Yeager et al. (2011)	Unnamed PBOP	USA

Studies from Germany examined the GIP, GESIS Panel, GESIS Online Panel Pilot (Struminskaya et al., 2014, 2016) and the Mannheim Corona Study (Cornesse et al., 2022b). Dutch studies focused on the LISS Panel (Leenheer & Scherpenzeel, 2013; Lugtig et al., 2014; Revilla, 2013; Scherpenzeel & Bethlehem, 2011) and a panel operated by Survey Sampling International LLC (Spijkerman et al., 2009). In Australia, research addressed Life in Australia (Kaczmirek et al., 2019; Kocar & Biddle, 2023), MyView (Hemsworth et al., 2021) and the ANU Poll (Pennay et al., 2018). South Korea was represented by studies on the KAMOS panel (Cho et al., 2017) and Gallup Korea's online panel (Seol et al., 2023). Finland was represented by a study on the eOpinion panel (Grönlund & Strandberg, 2014), Norway by the Norwegian Citizen Panel (Høgestøl & Skjervheim, 2014), Spain by a panel affiliated with PACIS recruitment (Arcos et al., 2020) and Singapore by the Singapore Life Panel (Vaithianathan et al., 2021). One cross-national study, CRONOS (Bottoni & Fitzgerald, 2021), included data from Estonia, Great Britain and Slovenia.

2.2 Data extraction and meta-analytic procedure

RB estimates for 1,897 items from 44 reports were included in the analyses. For each study, we recorded panel name, country, measured variables, measurement level (nominal, ordinal, or interval), domain, sensitivity and the operationalisation of bias. RB estimates were either extracted directly or calculated using the most straightforward method. All RB values were converted to absolute values to reflect the magnitude of bias regardless of direction. All references to RB in this study refer to absolute RB. Estimates were recorded at the level presented by the original authors. When both unweighted and weighted estimates were reported, only weighted estimates were coded. If multiple benchmarks were provided, the first reported benchmark was used. Variance was approximated as the inverse of panel size and analyses were based on the absolute value of RB.

Sensitivity was coded according to the guidelines from the Survey Quality Predictor (SQP; 2017). Variables were coded based on their potential for socially desirable responding: *Not present* (no potential), *A bit* (moderate potential) and *A lot* (highly sensitive topics, e.g., illegal behaviours or stigmatised attitudes). An example of this coding is provided in Table 2. Each variable was also assigned to a domain. Drawing on Saris and Gallhofer (2014), the SQP (2017) identifies eleven domains: National Politics, European Union Politics, International Politics, Family, Personal Relations, Work, Consumer Behaviour, Leisure Activities, Health, Living Conditions and Background Variables and Other Beliefs. For analysis, these were grouped into broader categories: *Politics* (International and National Politics), *Personal Relations* (Family and Personal Relations), *Miscellaneous* (Other Beliefs and Leisure Activities) and *Demographics* (Living Conditions and Background Variables, relabelled for clarity).

Table 2. *Example Survey Items for Each Sensitivity Level Following SQP (2017) Guidelines.*

Sensitivity Level	**Description**	**Typical topics**	**Example Item**
Not present	Item has no socially desirable or sensitive content.	Items related to demographics and typical behaviours.	Married
A bit	Item may elicit mild social desirability; some respondents might misreport.	Items related to personal finances, health status, charitable behavior, cultural activities, or evaluative judgments about institutions or individuals.	Household income $50K –59.9K
A lot	Item covers highly sensitive or stigmatized content, prone to misreporting.	Items related to racism, violence, religion, voting, crime, sexuality and drug use.	Ever tried cocaine

A random-effects approach was used, assuming the included studies represented a random sample from a broader population of relevant research (Raudenbush, 2009). Many studies contributed multiple effect sizes, as data quality measures were reported for several variables. Consequently, multiple RB estimates were extracted per study. To account for the dependency structure, a three-level meta-analytic model was applied (Harrer et al., 2021), modelling variance between studies (Level 3), between effect sizes within studies (Level 2) and sampling variance (Level 1). Heterogeneity (τ^2) at each level was estimated using the restricted maximum likelihood (REML) method (Viechtbauer, 2005) and the I^2 statistic (Higgins & Thompson, 2002) and Q-test for heterogeneity (Cochran, 1954) were reported. Sampling variances (Level 1) were treated as known and calculated as the inverse of the panel or sample size for each effect size (1/n). The model estimated overall RB and where significant heterogeneity was detected, moderator analyses were conducted to explain between- and within-study variance.

All models were estimated using restricted maximum likelihood in R (version 4.3.3; R Core Team, 2020) with the rma.mv() function from the metafor package (Viechtbauer, 2010). Model coefficients were tested using two-sided z-tests. Moderator significance (excluding the intercept) was assessed using the Wald-type QM test test, as implemented in rma.mv(). Reference categories for categorical predictors were set as follows: *USA* (Country), *Not present* (Sensitivity), *Nominal* (Level) and *Demographics* (Topic). Moderator variables were dummy-coded and tested both individually and jointly for significant predictors.

To assess the proportion of variance attributable to different sources of heterogeneity, I^2 values were calculated for Level 2 (within-study) and Level 3 (between-study) variance components, with total variance defined as the sum of all three levels. I^2 values reflected the percentage of total variance due to heterogeneity rather than sampling error. Pseudo R^2 values were computed to estimate the explanatory power of moderators by comparing total heterogeneity in the full model with that of a null model.

A sensitivity analysis was conducted by excluding the top 5% of effect sizes with the highest RB to examine the robustness of the results. The sensitivity analyses were conducted due to the presence of extremely large RB values in the full dataset, which, although methodologically valid, have the potential to influence the overall results. The distribution of RB values and the cutoff are illustrated in Figure 2. All meta-analytic models were re-estimated using this subset of the dataset.

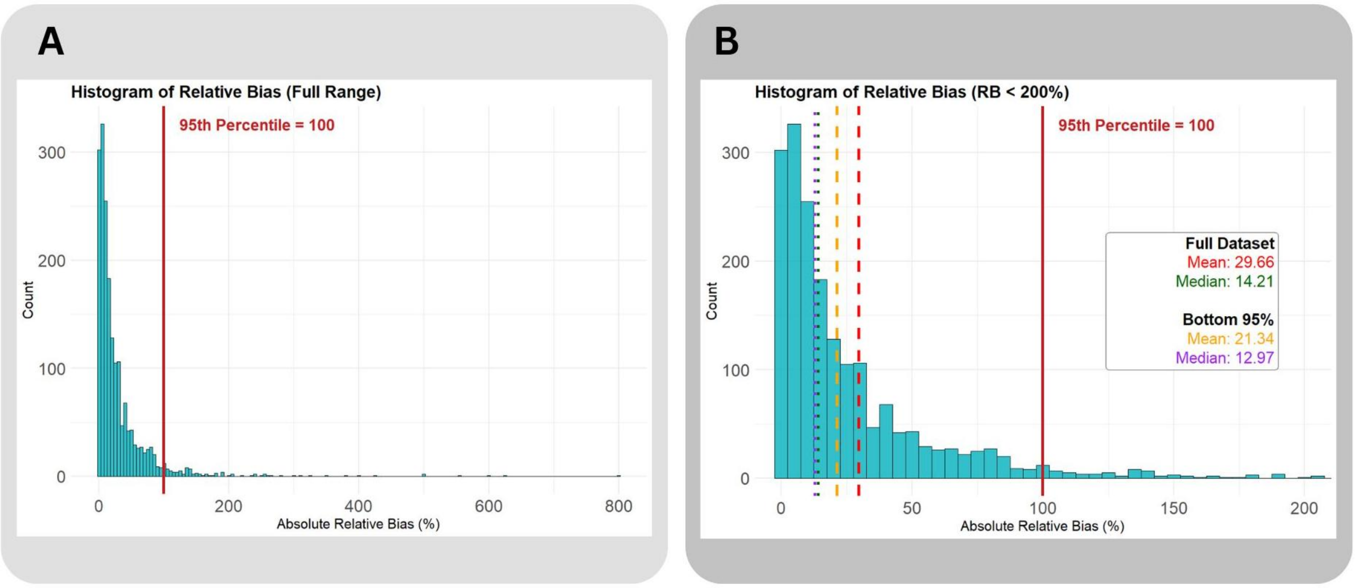

Figure 2. *Distribution of absolute Relative Bias (RB) Estimates. Panel A: Histogram showing the distribution of RB values across the full dataset (k = 1,897). The vertical red line indicates the 95th percentile cutoff (RB = 100%), used to define the upper limit for sensitivity analysis (k = 1,809). Panel B: Zoomed-in view of the same histogram, limited to RB values below 200%. Vertical dashed and dotted lines represent the mean and median RB, respectively, for both the full dataset and the bottom 95% subset. A summary box displays the corresponding statistics.*

3 RESULTS

3.1 Item characteristics

The most prevalent domain was *Living conditions and background variables* (44.3%), reflecting a strong emphasis on respondents' demographic and socio-economic characteristics (Table 3). *National politics* accounted for 23.7% and *Health* for 10.6% of all items. Other domains each contributed less than 10%. Due to the limited number of items in some SQP-defined domains (e.g., *Personal relations*), thematically similar domains were consolidated into broader categories. The distribution across these revised categories is shown in Table 4. *Demographics* remained the most common topic (44.3%), followed by *Politics* (25.4%), *Health* (10.6%) and *Personal relations* (8.33%). Although *Consumer behaviour*, *Work* and *Miscellaneous* topics were less frequent, each included at least 55 items.

Table 3. *Distribution of Survey Items by Domain.*

Domain	Number of items	Share of items
Living conditions and background variables	840	44.3%
National politics	450	23.7%
Health	202	10.6%
Family	142	7.49%
Consumer behaviour	90	4.74%
Work	55	2.9%
Other beliefs	49	2.58%
International politics	31	1.63%
Leisure activities	22	1.16%
Personal relations	16	0.84%

Table 4. *Distribution of Survey Items by Topic.*

Topic	Number of items	Share of items
Demographics (R)	840	44.3%
Politics	481	25.4%
Health	202	10.6%
Personal relations	158	8.33%
Consumer behaviour	90	4.74%
Miscellaneous	71	3.74%
Work	55	2.9%

Note: Categories preceded by the letter "R" will be used as the reference category in the meta-regression analysis that follows.

Items also varied by level of measurement (see Table 5). The majority were *nominal* (52.0%), followed by *ordinal* (45.7%). Only 43 items (2.3%) were measured at the *interval* level.

Table 5. *Distribution of Survey Items by Measurement Level.*

Level	Number of items	Share of items
Nominal (R)	987	52.0%
Ordinal	867	45.7%
Interval	43	2.27%

Note: Categories preceded by the letter "R" will be used as the reference category in the meta-regression analysis that follows.

According to the SQP (2017) coding instructions, 58.5% of items were classified as *Not present*, 34.5% as *A bit* and 7.0% as *A lot* (see Table 6) with regards to sensitivity.

Table 6. *Distribution of Survey Items by Sensitivity.*

Sensitivity	Number of items	Share of items
Not present (R)	1110	58.5%
A bit	654	34.5%
A lot	133	7.01%

Note: Categories preceded by the letter "R" will be used as the reference category in the meta-regression analysis that follows.

Most items and panels originated from the USA, which accounted for 63% of all items and contributed data from 13 panels (Table 7). Notably, five unnamed US panels were treated as separate entities, though some may overlap with named panels. Nonetheless, US-based items and panels were predominantly represented. The Netherlands (7.8%), Germany (6.64%) and Australia (4.96%) each contributed multiple panels and more than 94 items. Other countries—Finland, Norway, South Korea, Singapore, Spain, Estonia, Great Britain and Slovenia—were represented by one or two panels each, with Finland contributing 4.85% of items and the remainder contributing less.

Table 7. *Distribution of Panels and Survey Items by Country*

Country	Number of panels	Number of items	Share of items
USA (R)	13	1196	63.00%
The Netherlands	2	148	7.80%
Germany	4	126	6.64%
Australia	3	94	4.96%
Finland	1	92	4.85%
Norway	1	74	3.90%
South Korea	2	65	3.43%
Singapore	1	35	1.85%
Spain	1	22	1.16%
Estonia	1	15	0.79%
GB	1	15	0.79%
Slovenia	1	15	0.79%

Note: Categories preceded by the letter "R" will be used as the reference category in the meta-regression analysis that follows.

3.2 Meta-analysis

A total of $k = 1{,}897$ studies were included in the analysis. The estimated average standardised mean difference, based on a random-effects model, was $\hat{\mu} = 29.66$ (95% CI: 27.32 to 32.00), indicating a statistically significant deviation from zero ($z = 24.83$, $p < .0001$). The Q-test suggested substantial heterogeneity among true outcomes, Q(1896) = 22,292,792,423.81, $p < .0001$, with $\tau^2 = 2{,}707.19$ and $I^2 = 100.00\%$.

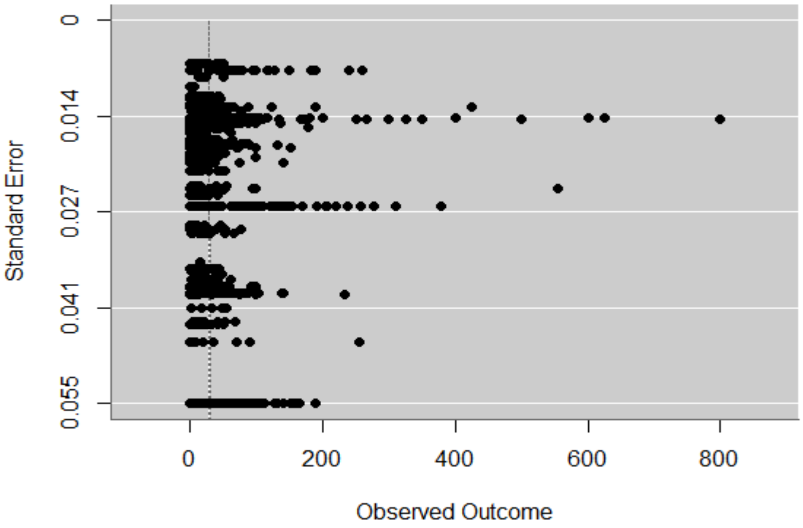

Figure 3. *Funnel Plot of RB Estimates (Full Sample, k = 1,897).*

A sensitivity analysis, excluding the top 5% most extreme RB estimates ($k = 1{,}809$), resulted in a slightly lower pooled RB of 21.34% (95% CI: 20.29% to 22.38%, $p < .001$). Funnel plots for the main analysis (Figure 3) and the sensitivity analysis (Figure 4) indicate that some studies reported exceptionally high levels of RB, which impacted the RB level. The influence of the extreme cases is also corroborated by the unweighted median finding, which was 12.21% (Figure 2).

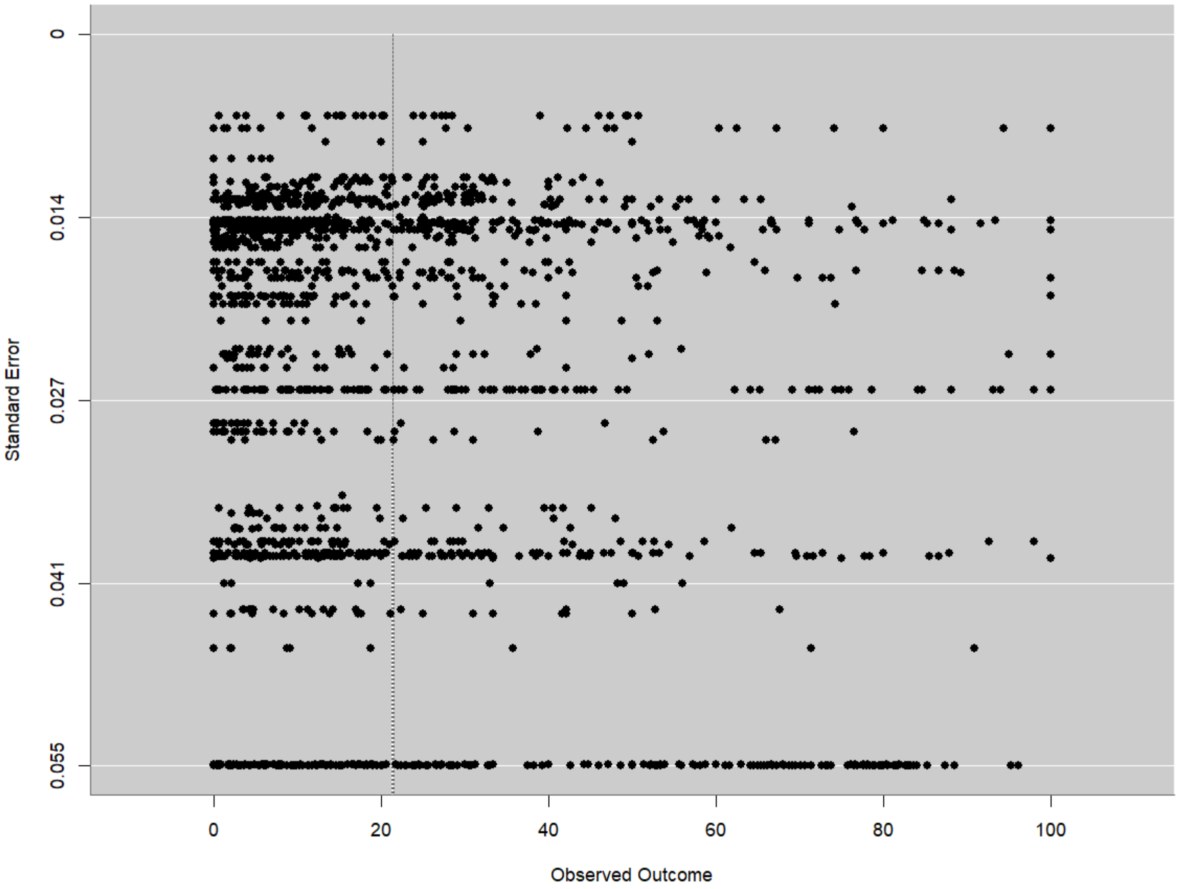

Figure 4. *Funnel Plot of RB Estimates (Sensitivity Analysis, Top 5% Excluded, k = 1,809).*

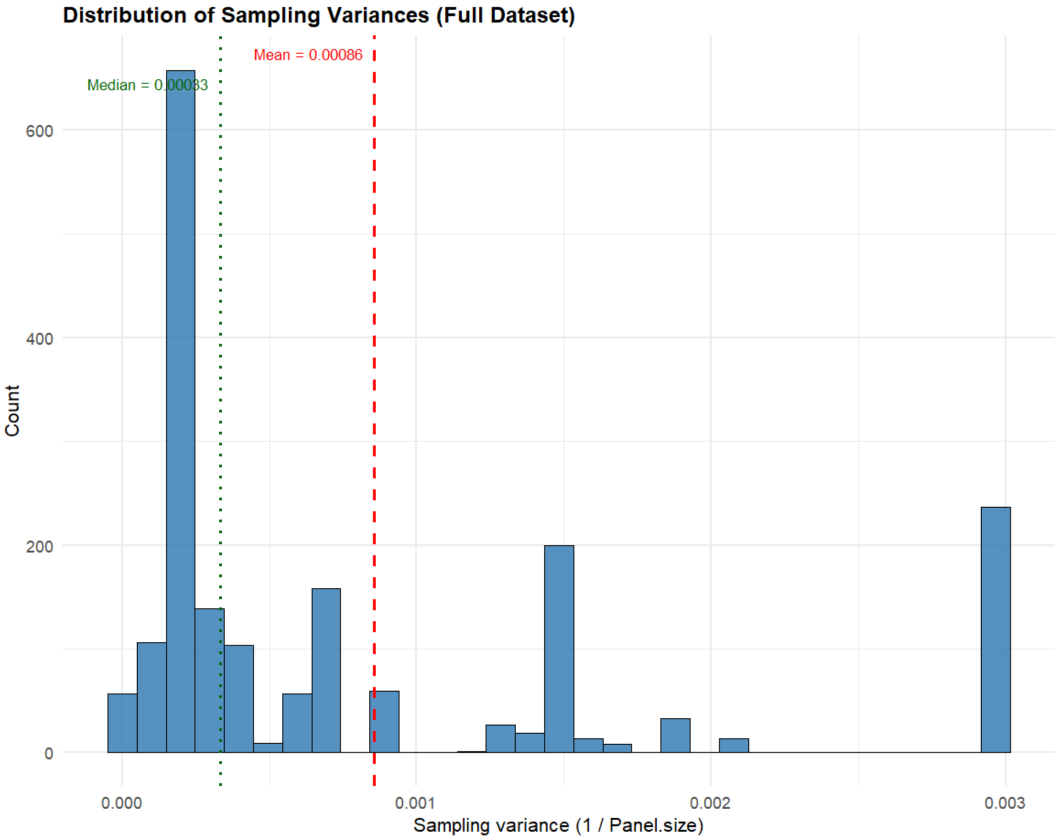

Figure 5. *Distribution of Sampling Variances in the Full Dataset.*

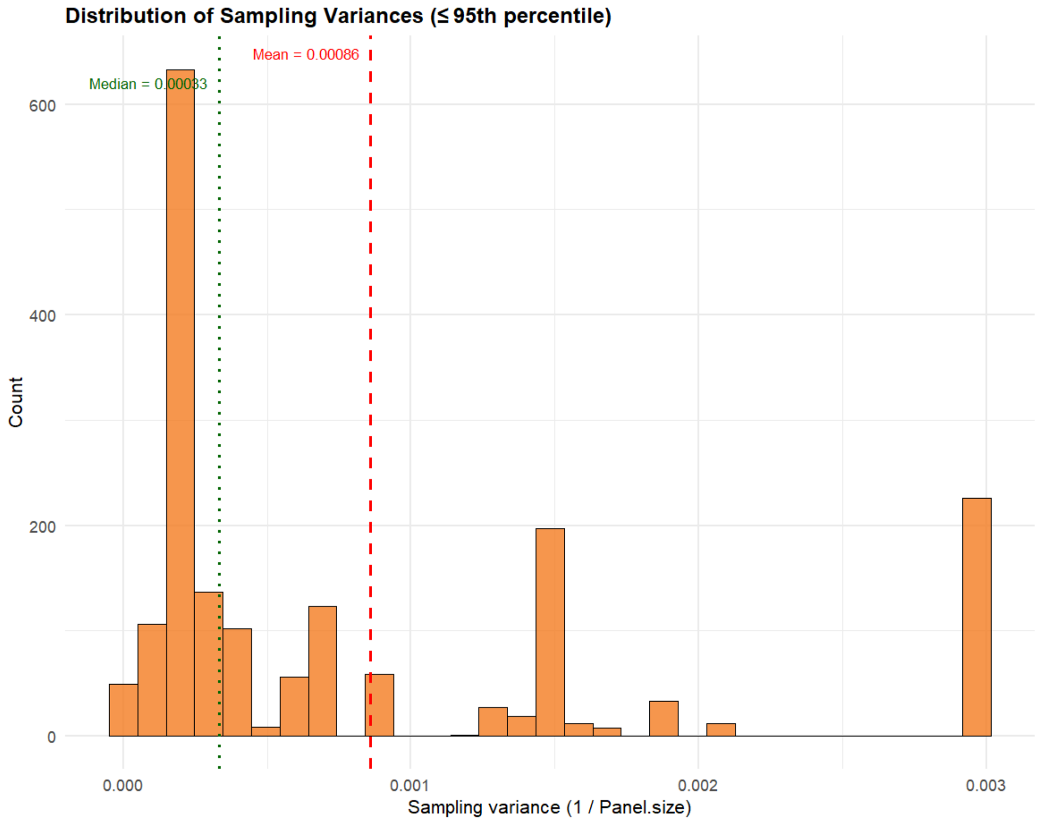

***Figure 6.** Distribution of Sampling Variances in the Full Dataset.*

As most panels featured very large sample sizes—and consequently very small sampling variances—the I^2 statistic was inflated towards 100% (Migliavaca et al., 2022). This pattern was consistent across all subsequent meta-analyses. Distributions of the approximated sampling variances for the full and sensitivity analyses are shown in Figure 5 and Figure 6, respectively.

3.3 Multi-level meta-analysis

The meta-analysis included $k = 44$ studies, from which 1,897 effect sizes were extracted. On average, 43.11 effect sizes were drawn per study (SD = 58.82; range = 4–316). Based on a three-level meta-analytic model, the pooled RB was 23.14% (95% CI [18.38%, 27.91%], $p < .001$). Estimated variance components were $\tau^2_{\text{Level 3}} = 157.31$ and $\tau^2_{\text{Level 2}} = 2{,}514.58$. Accordingly, $I^2_{\text{Level 3}} = 5.89\%$ of the total variance in RB was attributable to between-study differences, while $I^2_{\text{Level 2}} = 94.11\%$ was due to within-study differences (i.e., between effect sizes within the same study). Given this heterogeneity, testing moderators of the summary effect was warranted.

For the sensitivity analysis (Appendix B, Table B1), the meta-analysis was repeated excluding the top 5% most extreme RB values ($k = 1{,}809$). The pooled RB estimate decreased to 18.55% (95% CI [16.21%, 20.90%], $p < .001$). While the direction and statistical significance of the effect remained robust, the magnitude of the pooled estimate differed by 4.6% due to the excluded values.

3.4 Moderator analyses

The RB for the reference category, the U.S., based on the three-level meta-analytic model, was 23.93% (95% CI [16.81%, 31.06%], $p < .001$). Estimated variance components were $\tau^2_{\text{Level 3}} = 183.68$ and $\tau^2_{\text{Level 2}} = 2{,}517.96$ (Table 8), indicating that 6.80% of the total variance was due to between-study heterogeneity (Level 3) and 93.20% to within-study heterogeneity (Level 2).

A moderator analysis tested whether RB varied across countries. The omnibus test was not statistically significant, $Q_M\ (11) = 4.61$, $p = .948$, indicating no meaningful differences in RB across countries. No country-specific contrasts were statistically significant.

A sensitivity analysis excluding the top 5% most extreme RB values (k = 1,809) showed that the pooled RB estimate decreased to 17.84% (95% CI [14.40%, 21.27%]). Between-study heterogeneity increased slightly to 9.12%, but no significant country differences were observed (Appendix B, Table B1).

Table 8. *Effects of Country on RB.*

Moderator variable	**Q(df)**		***p***	**Level 2 variance**	**Level 3 variance**
Country	4.61 (11)		0.948	2517.96	183.68
Moderator levels	**# Studies**	**# ES**		**Intercept (95% CI)**	**β₁ (95% CI)**
USA (RC)	22	1196		23.93 (16.81; 31.06)***	
Australia	4	94		14.22 (–4.20; 32.63)	–9.72 (–28.13; 8.70)
Estonia	1	15		17.72 (–19.71; 55.15)	–6.21 (–43.65; 31.22)
Finland	1	92		33.31 (3.95; 52.46)	9.37 (–19.98; 38.72)
Great Britain	1	15		14.65 (–12.78; 52.00)	–9.28 (–46.71; 28.15)
Germany	6	126		15.96 (–0.26; 32.18)	–7.98 (–24.20; 8.25)
Norway	1	74		24.30 (–5.48; 54.09)	0.37 (–29.42; 30.15)
Singapore	1	35		15.26 (–8.88; 47.40)	–8.67 (–40.81; 23.46)
Slovenia	1	15		18.54 (–18.89; 55.90)	–5.39 (–42.82; 32.04)
South Korea	2	65		22.55 (–1.00; 47.87)	–1.39 (–24.92; 22.14)
Spain	1	22		27.32 (–7.27; 61.27)	3.39 (–31.20; 37.97)
The Netherlands	5	148		32.79 (16.08; 49.51)	8.86 (–7.85; 25.58)

Note: Asterisks indicate levels of statistical significance: * p < 0.05, ** p < 0.01, *** p < 0.001.

Based on the three-level meta-analytic model, the RB for the reference category—*nominal* measurement level—was 29.87% (95% CI [24.55%, 35.19%], p < .001). The estimated variance components were $\tau^2_{\text{Level 3}}$ = 161.36 and $\tau^2_{\text{Level 2}}$ = 2,469.39 (Table 9), indicating that 6.13% of the total variance was attributable to between-study heterogeneity and 93.87% to within-study heterogeneity.

A moderator analysis tested whether RB differed by measurement level. The omnibus test was statistically significant, Q_M (2) = 35.55, p < .001, indicating variation across levels. Follow-up contrasts showed that RB was significantly lower for studies using *ordinal* scales compared to *nominal* (β = –14.90, 95% CI [–19.81, –9.99], p < .001), while no significant difference was observed for *interval* scales (β = –9.82, 95% CI [–26.84, 7.21], p = .26).

In the sensitivity analysis excluding the top 5% most extreme RB values (k = 1,809; Appendix B, Table B2), the pattern of results remained. *Ordinal* scales continued to show significantly lower RB than *nominal* (β = –4.56, 95% CI [–6.75, –2.37], p < .001), while *interval* scales again showed no significant difference. However, the effect size for *ordinal* scales was reduced, with RB increasing from 14.98% to 25.31%.

Table 9. *Effects of measurement level on RB*

Moderator variable	**Q(df)**		***p***	**Level 2 variance**	**Level 3 variance**
Measurement Level	35.55 (2)		<0.001	2469.39	161.36
Moderator levels	**# Studies**	**# ES**		**Intercept (95% CI)**	**β₁ (95% CI)**
Nominal (R)	42	987		29.87 (24.55; 35.19)***	
Interval	9	43		20.06 (2.67; 37.15)	–9.82 (–26.84; 7.21)
Ordinal	38	867		14.98 (10.07; 19.89)	–14.90 (–19.81; –9.99)***

Note: Asterisks indicate levels of statistical significance: * p < 0.05, ** p < 0.01, *** p < 0.001.

The RB for the reference category of sensitivity, *Not present*, based on the three-level meta-analytic model, was 22.47% (95% CI [17.77%, 27.17%], p < .001). Variance components were $\tau^2_{\text{Level 3}}$ = 121.14 and $\tau^2_{\text{Level 2}}$ = 2,505.13 (Table 10), indicating that 4.61% of the total variance was attributable to between-study heterogeneity and 95.39% to within-study heterogeneity.

A moderator analysis tested whether RB varied by degree of sensitivity. The omnibus test for subgroup differences was statistically significant, Q_M (2) = 16.27, p = .0003. RB did not differ significantly between *Not present* and *A bit* (p = .895), but studies coded as *A lot* showed significantly higher RB than the reference category (β = 19.33, 95% CI [9.58, 29.08], p = .0001).

In the sensitivity analysis excluding the top 5% of extreme RB values (k = 1,809), *A lot* items still exhibited significantly higher RB (β = 9.84, 95% CI [5.34, 14.34], p < .001), though the effect size was reduced. RB for *A lot* items decreased from 41.8% to 32.31% (Appendix B, Table B3). Additionally, *A bit* items had lower RB (15.64%) than those coded as *Not present* (19.01%).

Table 10. *Effects of sensitivity on RB.*

Moderator variable	**Q(df)**	***p***	**Level 2 variance**	**Level 3 variance**
Sensitivity	16.27 (2)	0.0003	2505.13	121.14
Moderator levels	**# Studies**	**# ES**	**Intercept (95% CI)**	**β_1 (95% CI)**
Not present (R)	42	1110	22.47 (17.77; 27.17)***	
A bit	33	654	22.10 (16.46; 27.74)	–0.38 (–5.96; 5.20)
A lot	11	133	41.8 (32.81; 50.87)	19.33 (9.58; 29.08)***

Note: Asterisks indicate levels of statistical significance: * p < 0.05, ** p < 0.01, *** p < 0.001.

The RB for the reference category, *Demographics*, based on the three-level meta-analytic model, was 21.94% (95% CI [16.79%, 27.09%], p < .001). Estimated variance components were $\tau^2_{\text{Level 3}}$ = 132.77 and $\tau^2_{\text{Level 2}}$ = 2,515.29 (Table 11), indicating that 5.01% of the total variance was due to between-study heterogeneity (Level 3) and 94.99% to within-study heterogeneity (Level 2).

A moderator analysis tested whether RB varied by topic. The omnibus test was not statistically significant, Q_M(6) = 10.05, p = .123, indicating no meaningful differences in RB across topics. None of the topic-specific contrasts reached statistical significance. A sensitivity analysis excluding the top 5% of RB values yielded similar results, with topic again not identified as a significant moderator (Appendix B, Table B4).

Table 11. *Effects of topic on RB.*

Moderator variable	**Q(df)**	***p***	**Level 2 variance**	**Level 3 variance**
Topic	10.05 (6)	.123	2515.29	132.77
Moderator levels	**# Studies**	**# ES**	**Intercept (95% CI)**	**β_1 (95% CI)**
Demographics (R)	38	840	21.94 (16.79; 27.09)***	
Consumer behaviour	12	90	28.26 (18.95; 37.57)	6.32 (–5.66; 18.30)
Health	19	202	27.78 (20.08; 35.47)	5.85 (–3.03; 14.72)
Miscellaneous	13	71	31.86 (18.88; 44.83)	9.92 (–3.41; 23.25)
Personal relations	23	158	15.36 (5.92; 24.81)	–6.58 (–16.01; 2.85)
Politics	14	481	28.64 (20.47; 36.80)	6.70 (–1.47; 14.87)
Work	17	55	20.73 (8.80; 32.66)	–1.21 (–15.35; 12.94)

Note: Asterisks indicate levels of statistical significance: * p < 0.05, ** p < 0.01, *** p < 0.001.

Including both *sensitivity* and *measurement level* as moderators, the RB for the reference category—*Not present* sensitivity and *nominal* measurement level—was 28.23% (95% CI [22.94%, 33.51%], p < .001) based on the three-level meta-analytic model. Variance components were estimated as $\tau^2_{\text{Level 3}}$ = 133.29 and $\tau^2_{\text{Level 2}}$ = 2,467.31 (Table 12), indicating that 5.13% of total variance was due to between-study heterogeneity and 94.87% to within-study heterogeneity.

Moderator analysis tested whether RB varied by *sensitivity* and *measurement level*. The omnibus test was significant, Q_M (4) = 44.33, $p < .001$, indicating the moderators jointly explained variation in RB. Contrasts revealed significantly higher RB for variables with *A lot* of sensitivity compared to the reference category ($\beta = 14.88$, 95% CI [5.06, 24.71], $p = .003$), while no significant difference was found for *A bit* ($p = .512$). For *measurement level*, RB was significantly lower for *ordinal* than *nominal* variables ($\beta = –13.69$, 95% CI [–18.73, –8.65], $p < .001$), while the contrast for *interval* variables was not significant ($p = .287$).

Results from the sensitivity analysis (excluding the top 5% of extreme RB values; k = 1,809) were largely consistent (Appendix B, Table B5). Both *A lot* sensitivity and *ordinal* measurement remained significantly associated with RB. However, effect sizes were smaller: *nominal* variables with *A lot* of sensitivity had an RB of 29.05% and *ordinal* variables with *Not present* sensitivity had an RB of 17.01%. Additionally, a significant effect was observed for *nominal* variables with *A bit* of sensitivity, with an RB of 17.34%.

Table 12. *Effects of sensitivity and measurement level on RB.*

Moderator variable	**Q(df)**		***p***	**Level 2 variance**	**Level 3 variance**
Sensitivity and Level	44.33 (4)		< 0.0001	2467.31	133.29
Moderator levels	**# Studies**	**# ES**	**Intercept (95% CI)**		**β_1 (95% CI)**
Not present, Nominal (R)	40	641	28.23 (22.94; 33.51) ***		—
A bit, Nominal	22	227	30.11 (24.32; 35.90)		1.88 (–3.74; 7.51)
A lot, Nominal	10	119	43.11 (34.00; 52.23)		14.88 (5.06; 24.71) **
Not present, Interval	4	25	19.06 (6.20; 31.91)		–9.17 (–26.04; 7.69)
Not present, Ordinal	34	444	14.54 (9.50; 19.57)		–13.69 (–18.73; –8.65) ***

Note: Asterisks indicate levels of statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

However, the test for residual heterogeneity remained significant, Q_E (df = 1,892) = 2.03 × 10^{10}, $p < .0001$, indicating substantial unexplained variability across effect sizes despite the inclusion of moderators. To assess the explanatory power of *measurement level* and *sensitivity*, we examined the proportion of between-effect heterogeneity accounted for by each moderator individually and jointly. Table 13 reports the variance components at the article level ($\tau^2_{\text{Level 3}}$) and the within-article effect-size level ($\tau^2_{\text{Level 2}}$), the corresponding proportions of variance (I^2) and the proportion of total heterogeneity explained (pseudo R^2), calculated as the proportional reduction in total between-effect variance ($\tau^2_{\text{Level 3}} + \tau^2_{\text{Level 2}}$) relative to the null model. In the null model, $I^2_{\text{Level 2}}$ was 94.11%, indicating that the vast majority of variance occurred between effect sizes within the same study, rather than between studies ($I^2_{\text{Level 3}}$ = 5.89%). Including measurement level slightly increased the proportion of variance attributed to between-study differences ($I^2_{\text{Level 3}}$ = 6.13%), suggesting this moderator introduced some structure at the study level, though the change was minimal. Similarly, sensitivity reduced $I^2_{\text{Level 3}}$ to 4.61%, implying it may be more relevant for explaining within-study variance. When both moderators were included, $I^2_{\text{Level 3}}$ increased slightly to 5.13% and $I^2_{\text{Level 2}}$ decreased slightly to 94.87%, indicating only a modest improvement in model fit. According to the pseudo R^2, including *measurement level* alone explained approximately 1.54% of the total heterogeneity; *sensitivity* explained slightly more (1.71%). When both were included, they jointly accounted for 2.67% of the total heterogeneity. In the 95% subset used for the sensitivity analysis, the combined explanatory power was 2.44% (Appendix B, Table B6).

Table 13. *Variance Components and Explained Heterogeneity of the Meta-analytic Models.*

Model	**τ^2 Level 3**	**τ^2 Level 2**	**Total τ^2**	**I^2 Level 3 (%)**	**I^2 Level 2 (%)**	**Pseudo R^2 Total (%)**
Null multi-level model	157.31	2514.576	2671.887	5.89	94.11	–
Level only	161.361	2469.391	2630.752	6.13	93.87	1.54
Sensitivity only	121.143	2505.126	2626.269	4.61	95.39	1.71
Level and Sensitivity	133.288	2467.313	2600.601	5.13	94.87	2.67

4 DISCUSSION

The current meta-analysis synthesised data from 44 studies, comprising 1,897 effect sizes. The pooled estimate of RB was 23.14%, indicating a moderately high level of bias. In relation to RQ1, this finding suggests that, despite rigorous design and probability-based sampling, PBOPs do not fully eliminate bias relative to external benchmarks. A multilevel model showed that most variability was attributable to within-study heterogeneity, with only 5.89% due to between-study differences. This indicates that item characteristics contribute more to variation in *RB* than panel characteristics.

Although country was hypothesised to moderate RB due to cultural and infrastructural differences, no significant variation was found across the 12 countries analysed. Addressing RQ2, this suggests that national-level traits may not systematically influence response bias in PBOPs. This aligns with prior research showing that comparable levels of Internet diffusion do not guarantee equivalence in measurement constructs across contexts (Büchi et al., 2016). In other words, cultural or infrastructural differences may not systematically affect response biases in standardised instruments. As Huijsmans et al. (2021) demonstrate, much of the variation in political attitudes exists below the national or municipal level, indicating that broader contextual factors explain little of the variance in attitudinal outcomes.

We also tested the survey item *topic* as a potential moderator and found some variation in RB across topics, but none were statistically significant compared to Demographics, which had a baseline RB of 21.94%. Addressing RQ2, this suggests that broad topic domains may not strongly influence RB in PBOPs. This contrasts with the Survey Quality Predictor (Felderer et al., 2024), which identified topic-related characteristics as key predictors of reliability and validity. In our analysis, other item characteristics, such as sensitivity and measurement level, were more influential in predicting RB. Similarly, Groves and Peytcheva (2008), in a meta-analysis of nonresponse bias, found that the topic was not consistently related to bias across hundreds of estimates. They argued that broad topic categories may be too general to explain variation in response behaviour, with most variation occurring at the level of individual questions rather than across topics.

Measurement level emerged as a significant moderator. Addressing RQ2, RB was significantly lower for items measured on *ordinal* scales than on *nominal* ones. Classifying variables as *nominal* or low-category *ordinal* can lead to information loss, reduced statistical power and biased estimates, particularly for non-normal traits (Verhulst & Neale, 2022). *Nominal* measures provide only basic classification and limited statistical utility (Idika et al., 2023), whereas *ordinal* outcomes retain more information and offer greater power than dichotomised or *nominal* formats (Selman et al., 2023). Simulation studies further show that misclassifying latent continuous variables inflates error rates and underestimates effect sizes (Liddell & Kruschke, 2018). Although *ordinal* measures offer greater precision, collapsing *interval* variables into *nominal* categories for benchmark comparisons discards meaningful variation, increases error, and may inflate response bias, helping explain the lower bias in ordinal items.

Items with *a lot* of sensitivity showed significantly higher RB compared to those with *no* sensitivity. Items coded as *a bit* sensitive did not differ significantly from non-sensitive items. These findings, related to RQ2, suggest that only high sensitivity meaningfully impacts response bias. However, in the sensitivity analysis, *a bit* sensitive items showed 3.37% less RB than non-sensitive items. The finding that highly sensitive items exhibit greater RB aligns with evidence that sensitive topics are more susceptible to response distortion (Tourangeau & Yan, 2007; Nayak & Narayan, 2019). Tourangeau & Yan (2007) also found that social desirability pressure predicts misreporting in surveys on sensitive topics. The lower bias for moderately sensitive items may be explained by survey mode differences: benchmarks were typically interviewer-administered, while PBOPs used self-completion online formats, which can reduce social desirability bias (Berzelak & Vehovar, 2018). For highly sensitive topics, however, rapport in face-to-face settings may encourage more honest reporting than anonymous online modes (Westland et al., 2024). Overall, while high sensitivity reliably increases RB, its effects may interact in complex ways with other survey and item characteristics.

When tested jointly, sensitivity and measurement level each retained a significant effect. Again, in relation to RQ2, this shows that both characteristics independently contribute to variation in RB. High sensitivity (*A lot*) remained associated with higher RB, while *ordinal* measurement continued to show lower RB. However, together these moderators explained only 2.67% of the total heterogeneity. This modest explanatory power and the substantial residual heterogeneity suggest that unmeasured factors, such as survey design, also play a role. For example, whether the offline population is included and whether the survey is administered in a unified or mixed-mode format, can affect data quality. Panels that provide electronic devices and internet access may promote greater measurement equivalence than those using mixed modes such as paper questionnaires (Blom et al., 2015). Other relevant features include panel versus cross-sectional design, recruitment strategy (e.g., fresh sample vs. follow-up) and how offline participants are treated (Bosch & Maslovskaya, 2023). Differences in recruitment approaches and respondent motivation may also influence satisficing behaviours such as straight-lining or nonresponse (Cornesse & Blom, 2023). Factors like incentives and recruitment methods can affect response quality beyond the influence of sensitivity or measurement level. Moreover, variation may arise at additional levels beyond the panel and effect size. Some studies, for instance, evaluate data quality across multiple samples within a PBOP under varying experimental conditions (e.g., different question wordings in Smith et al., 2003). Characteristics at the study, sample and variable levels may further contribute to heterogeneity. Future research would benefit from a multilevel meta-analytic approach incorporating a broader range of moderators to capture these multiple sources of variation.

The sensitivity analysis, detailed in Appendix B, did not fully corroborate the main findings. While the general direction of effects remained consistent, the magnitude of those effects was notably reduced. This change is likely due to a small number of highly influential cases that were identified. Their exclusion suggests that certain extreme values contributed substantially to the observed heterogeneity.

Thus, we conclude that the development of ICT and the growing demand for faster, cost-effective data collection have led to the increasing popularity of online panels. While nonprobability panels are cheaper, they carry substantial bias (Sakshaug et al., 2019), and even probability-based panels, though more robust, do not eliminate errors and must be used with caution. Our meta-analysis found a pooled RB of 23.14%, exceeding common social research thresholds (5–10%) for concern, suggesting PBOPs may be unsuitable when precision and representativeness are critical. We may add that one of the key contributors to these biases is the low overall response rate in PBOPs. In fact, they typically achieve rates below 20% (Kocar & Kaczmirek, 2023), which is substantially lower than those observed in traditional probability-based face-to-face surveys, where rates of 50–60% are still common in large-scale academic studies (Vehovar & Beullens, 2017). Such low participation rates not only reduce the effective sample size but also heighten the risk of nonresponse bias, especially when nonrespondents differ systematically from respondents.

While the shift towards ICT-driven, interviewer-free data collection offers clear advantages in terms of cost reduction and operational efficiency, it appears that, in this case, technological automation may have progressed while methodological rigor was left behind. This is not unique to PBOPs—similar concerns have emerged in other domains, such as automated content analysis, predictive policing and algorithmic decision-making, where risks of bias, opacity and reduced accountability often accompany efficiency gains. For instance, algorithmic systems used in data analytics and AI-driven decision-making have been shown to produce socially biased outcomes, which may not stem directly from the technology itself, but from how it is designed, implemented and perceived in specific social and organisational contexts (Kordzadeh & Ghasemaghaei, 2021). The quest for efficiency in survey research must be balanced against the fundamental need for data quality. This trade-off between speed, costs and accuracy raises broader concerns about the limits of ICT-based automation in social research and invites reconsideration of when and how human involvement remains essential to ensure data integrity. In the broader context, the methodological weaknesses seen in PBOP implementation reflect broader issues in digital research practices, where the promise of scale and efficiency often overshadows the foundational need for accuracy and validity.

This trade-off is especially evident in PBOP survey instrument design, where item construction—particularly sensitivity and measurement scale—is critical. In some cases, the bias was extreme (the maximum identified was 800% for respondents who live on a boat, RV, van, etc.) indicating that certain items may be fundamentally unsuited to online survey modes. This reflects a broader issue of inadequate planning. Researchers often include items on low-prevalence behaviours without accounting for the required sample size or statistical power. In these contexts, even small misclassifications can cause large overestimations, and standard power calculations may be invalid when prevalence is low (Williams et al., 2007). Effective survey design should avoid items likely to yield highly biased estimates due to nonresponse or noncoverage and instead incorporate careful planning, including power analyses tailored to rare outcomes. In this context, we recommend using ordinal rather than nominal scales when feasible. Guidance on item sensitivity is more complex: PBOPs may reduce bias for some sensitive topics because self-administered online formats lessen social desirability pressures compared to interviewer-administered modes

(Berzelak & Vehovar, 2018; Lozar Manfreda et al., 2002), yet highly sensitive items may still require adjustment or avoidance.

Several limitations should be considered when interpreting these findings. First, sampling variance was approximated using the inverse of sample size, treating benchmark estimates as "true" values for calculating RB. However, benchmarks, which are often derived from large-scale governmental or official surveys, are not free from error. Interviewer-administered benchmarks may be affected by mode effects, interviewer bias, or question-order effects (Callegaro et al., 2015a; Schwarz et al., 2008). Additionally, benchmarks may differ subtly but meaningfully from the PBOP items they are compared to, in phrasing, time frames, or contextual cues, potentially inflating observed bias (Rasinski et al., 2012). In some instances, PBOP estimates may even be more accurate than benchmarks, particularly when benchmarks rely on outdated methods or offer lower respondent anonymity (Bialik, 2018). Benchmark quality may also differ, as we did not account for such variation and considered all official statistics or surveys using traditional methods as valid benchmarks. Some of these are large-scale governmental surveys aiming for high response rates, while others may be telephone surveys with lower response rates. Future analyses should consider accounting for these differences. As with other meta-analyses based on observational studies, there is a risk of bias due to unobserved or unmeasured confounders and the absence of key moderators that could not be included in the analysis. This is further compounded by potential non-reporting bias, where the availability of results may depend on their statistical significance or direction, leading to systematic differences between reported and unreported findings and threatening the validity of the synthesis (Page et al., 2024).

5 CONCLUSION

This meta-analysis demonstrates that, while PBOPs are methodologically robust, they do not eliminate response bias and cannot fully replace traditional probability-based surveys. Although they may perform well for many items, especially when those items are carefully designed, substantial bias persists for others. This is especially problematic for highly sensitive or low-prevalence measures, which may be fundamentally incompatible with the PBOP format. More broadly, this highlights a limitation in applying ICTs to survey research: despite their advantages, they are not universally reliable. More specifically, the low overall response rates are a particularly critical limitation of PBOPs. The flexibility of digital technologies can produce unintended consequences, raising challenges for responsible innovation (Stahl, 2017). Online surveys must therefore be designed with these limitations in mind. For instance, it is generally inadvisable to include items expected to elicit responses from only a small subset of participants; such items are better suited to specialised low-prevalence methodologies. Moreover, the successful implementation of ICTs requires more than technical infrastructure; it also demands cultural and methodological adaptation, which often remains underdeveloped (Bryda & Costa, 2023). Recognising that current PBOP methodologies may be appropriate for some items but not others offers a more realistic perspective. These findings are especially relevant for national statistical agencies and institutions seeking to replace or supplement traditional modes with PBOPs. Awareness of the identified biases and risks can inform instrument design and panel recruitment strategies. Continued innovation in data collection, rigorous study design and ongoing evaluation are essential for improving the quality of PBOP estimates.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding: This work was funded by the Slovenian Research Agency (grant numbers P5-0399, J5-3100 and J5-50159).

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: A.I.: Data Curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation. M.B.: Methodology, Supervision, Writing – Review & Editing. V.V.: Methodology, Supervision, Writing – Review & Editing.

Statement on the Use of Artificial Intelligence Tools: The authors declared that they used ChatGPT (an AI language model developed by OpenAI) solely for the purpose of proofreading and improving the clarity of the text in this article.

Data Availability: The data that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.15583040.

APPENDIX A: GLOSSARY OF KEY TERMS AND ABBREVIATIONS

Table A1. *Definitions of key terms and abbreviations.*

Term	**Definition**
PBOP (Probability-Based Online Panel)	An online survey panel in which participants are recruited using traditional probability sampling methods to ensure representativeness, with data collection conducted primarily online.
Nonprobability Online Panel	An online survey panel in which participants are recruited using nonprobability methods, such as advertising, convenience sampling, or volunteer sign-ups, without a known probability of selection.
Benchmark	A trusted reference dataset, typically from a high-quality source such as a government survey, official statistics, or other large-scale traditional survey, used as a standard for evaluating the accuracy of other survey estimates.
RB (Relative Bias)	A metric expressing the percentage difference between a survey estimate and a benchmark as a proportion of the benchmark value, often reported in absolute form to assess accuracy across variables and populations.

APPENDIX B: SENSITIVITY ANALYSES

Table B1. *Effects of country on RB (sensitivity analysis, top 5% excluded).*

Moderator variable	**Q(df)**	***p***	**Level 2 variance**	**Level 3 variance**
Country	8.12 (11)	.703	464.01	46.55
Moderator levels	**# Studies**	**# ES**	**Intercept (95% CI)**	**β_1 (95% CI)**
USA (R)	22	1124	17.84 (14.40; 21.27) ***	
Australia	4	94	14.10 (5.31; 22.89)	–3.74 (–12.53; 5.05)
Estonia	1	15	17.72 (–3.31; 38.10)	–0.12 (–17.71; 17.47)
Finland	1	87	25.21 (7.25; 43.18)	7.38 (–7.15; 21.91)
Great Britain	1	15	14.65 (–6.38; 35.02)	–3.19 (–20.78; 14.40)
Germany	6	126	15.89 (8.17; 23.60)	–1.95 (–9.66; 5.77)
Norway	1	73	22.78 (4.68; 40.88)	4.94 (–9.72; 19.60)
Singapore	1	34	11.59 (–7.44; 30.63)	–6.25 (–21.84; 9.34)
Slovenia	1	15	18.54 (–2.49; 39.80)	0.70 (–16.89; 18.30)
South Korea	2	64	20.67 (5.86; 35.47)	2.83 (–8.54; 14.20)
Spain	1	22	27.32 (7.40; 47.24)	9.48 (–7.00; 25.96)
The Netherlands	5	140	24.59 (13.08; 36.10)	6.75 (–1.33; 14.83)

Note: Asterisks indicate levels of statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B2. *Effects of measurement level on RB (sensitivity analysis, top 5% excluded).*

Moderator variable	Q(df)	*p*	Level 2 variance	Level 3 variance
Measurement Level	16.89 (2)	.0002	459.72	44.67
Moderator levels	**# Studies**	**# ES**	**Intercept (95% CI)**	**β_1 (95% CI)**
Nominal (R)	42	924	20.48 (17.87; 23.08) ***	
Interval	9	43	20.80 (13.61; 27.97)	0.32 (–7.24; 7.89)
Ordinal	38	842	15.92 (13.72; 18.12) ***	–4.56 (–6.75; –2.37) ***

Note: Asterisks indicate levels of statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B3. *Effects of sensitivity on RB (sensitivity analysis, top 5% excluded).*

Moderator variable	**Q(df)**	***p***	**Level 2 variance**	**Level 3 variance**
Sensitivity	32.70 (2)	<.001	457.02	37.56
Moderator levels	**# Studies**	**# ES**	**Intercept (95% CI)**	**β_1 (95% CI)**
Not present (R)	42	1,079	19.01 (16.66; 21.37) ***	
A bit	32	619	15.65 (13.24; 18.11) **	–3.37 (–5.84; –0.90) **
A lot	11	111	28.85 (22.00; 35.70) ***	9.84 (5.34; 14.34) ***

Note: Asterisks indicate levels of statistical significance: * p < 0.05, ** p < 0.01, *** p < 0.001.

Table B4. *Effects of topic on RB (sensitivity analysis, top 5% excluded).*

Moderator variable	**Q(df)**	***p***	**Level 2 variance**	**Level 3 variance**
Topic	4.64 (6)	.591	464.34	42.24
Moderator levels	**# Studies**	**# ES**	**Intercept (95% CI)**	**β_1 (95% CI)**
Demographics (R)	38	829	18.16 (15.58; 20.74) ***	
Consumer behaviour	12	87	17.26 (12.47; 22.54)	–0.90 (–6.19; 4.38)
Health	19	182	20.26 (13.66; 26.87)	2.10 (–1.93; 6.12)
Miscellaneous	13	65	23.43 (14.84; 32.03)	5.27 (–0.75; 11.28)
Personal relations	22	154	18.09 (13.42; 22.77)	–0.08 (–4.25; 4.09)
Politics	14	440	18.30 (14.15; 22.45)	0.14 (–3.62; 3.89)
Work	17	52	18.70 (11.85; 25.55)	0.53 (–5.73; 6.80)

Note: Asterisks indicate levels of statistical significance: * p < 0.05, ** p < 0.01, *** p < 0.001.

Table B5. *Effects of sensitivity and measurement level on RB (sensitivity analysis, top 5% excluded).*

Moderator variable	**Q(df)**	***p***	**Level 2 variance**	**Level 3 variance**
Sensitivity and Level	41.17 (4)	<.001	455.02	39.36
Moderator levels	**# Studies**	**# ES**	**Intercept (95% CI)**	**β_1 (95% CI)**
Not present, Nominal (R)	40	615	20.22 (17.63; 22.80) ***	
A bit, Nominal	21	211	17.34 (14.77; 19.92) *	–2.88 (–5.39; –0.38) *
A lot, Nominal	10	98	29.05 (21.92; 36.18) ***	8.83 (4.29; 13.38) ***
Not present, Ordinal	34	439	17.01 (14.58; 19.45) **	–3.21 (–5.46; –0.96) **
Not present, Interval	4	25	21.87 (13.86; 29.88)	1.65 (–5.85; 9.14)

Note: Asterisks indicate levels of statistical significance: * p < 0.05, ** p < 0.01, *** p < 0.001.

Table B6. *Variance Components and Explained Heterogeneity of the Meta-analytic Models (sensitivity analysis, top 5% excluded).*

Model	**τ^2 Level 3**	**τ^2 Level 2**	**Total τ^2**	**I^2 Level 3 (%)**	**I^2 Level 2 (%)**	**Pseudo R^2 Total (%)**
Null multi-level model	42.898	463.869	506.768	8.47	91.53	–
Level only	44.671	459.718	504.389	8.86	91.14	0.47
Sensitivity only	37.555	457.018	494.574	7.59	92.41	2.41
Level and Sensitivity	39.362	455.022	494.384	7.96	92.04	2.44

REFERENCES

References marked with an asterisk (*) were included among the studies analysed in the meta-analysis.

*<strong>Arcos, A., Rueda, M. d. M., & Pasadas-del-Amo, S.</strong> (2020). Treating nonresponse in probability-based online panels through calibration: Empirical evidence from a survey of political decision-making procedures. *Mathematics*, 8(3), 423. https://doi.org/10.3390/math8030423

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., Dillman, D. A., Frankel, M. R., Garland, P., Groves, R. M., Kennedy, C., Krosnick, J. A., Lavrakas, P. J., Lee, S., Link, M. W., Piekarski, L. B., Rao, K. N., Thomas, R. K., & Zahs, D. A. (2010). AAPOR report on online panels. *Public Opinion Quarterly*, 74, 711–781. https://doi.org/10.1093/poq/nfq048

*<strong>Bell, J., Huber, J., & Viscusi, W. K.</strong> (2011). Survey mode effects on valuation of environmental goods. *International Journal of Environmental Research and Public Health*, 8(4), 1222–1243. https://doi.org/10.3390/ijerph8041222

*<strong>Berrens, R. P., Bohara, A. K., Jenkins-Smith, H. C., Silva, C. L., & Weimer, D. L.</strong> (2003). The advent of internet surveys for political research: A comparison of telephone and internet samples. *Political Analysis*, 11, 1–22. https://doi.org/10.1093/pan/11.1.1

Berzelak, J., & Vehovar, V. (2009). Information technology in survey research. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (pp. 2024–2029). IGI Global. https://doi.org/10.4018/978-1-60566-026-4.ch318

Berzelak, N., & Vehovar, V. (2018). Mode effects on socially desirable responding in web surveys compared to face-to-face and telephone surveys. *Advances in Methodology and Statistics, 15*(2), 21–43. https://doi.org/10.51936/lrkv4884

Berzelak, N., Vehovar, V., & Manfreda, K.L. (2025). Nonresponse in Web Surveys. In Lovric, M. (ed*) International Encyclopedia of Statistical Science*. Springer. https://doi.org/10.1007/978-3-662-69359-9_429

Bialik, K. (2018, December 6). How asking about your sleep, smoking or yoga habits can help pollsters verify their findings. *Pew Research Center*. https://www.pewresearch.org/short-reads/2018/12/06/how-asking-about-your-sleep-smoking-or-yoga-habits-can-help-pollsters-verify-their-findings/

*<strong>Bilgen, I., Dennis, M. J., & Liebert, L.</strong> (2018). Nonresponse follow-up impact on AmeriSpeak panel sample composition and representativeness. *NORC*. https://amerispeak.norc.org/content/dam/amerispeak/research/pdf/Bilgen_etal_WhitePaper1_NRFU_SampleComposition.pdf

Blom, A. G., Bosnjak, M., Cornilleau, A., Cousteaux, A.-S., Das, M., Douhou, S., & Krieger, U. (2016). A comparison of four probability-based online and mixed-mode panels in Europe. *Social Science Computer Review*, 34(1), 8–25. https://doi.org/10.1177/0894439315574825

*<strong>Blom, A. G., Gathmann, C., & Krieger, U.</strong> (2015). Setting up an online panel representative of the general population: The German Internet Panel. *Field Methods*, 27(4), 391–408. https://doi.org/10.1177/1525822X15574494

*<strong>Blom, A. G., Herzing, J. M. E., Cornesse, C., Sakshaug, J. W., Krieger, U., & Bossert, D.</strong> (2017). Does the recruitment of offline households increase the sample representativeness of probability-based online panels? Evidence from the German Internet Panel. *Social Science Computer Review*, 35(4), 498–520. https://doi.org/10.1177/0894439316651584

Boland, M., Sweeney, M. R., Scallan, E., Harrington, M., & Staines, A. (2006). Emerging advantages and drawbacks of telephone surveying in public health research in Ireland and the U.K. *BMC Public Health*, 6, Article 208. https://doi.org/10.1186/1471-2458-6-208

Bosch, O. J., & Maslovskaya, O. (2023). GenPopWeb2: The utility of probability-based online surveys – Literature review. NCRM. https://www.ncrm.ac.uk/documents/GenPopWeb2_The%20utility%20of%20probability-based%20online%20surveys_Literature%20review.pdf

Bosnjak, M., Das, M., & Lynn, P. (2016). Methods for Probability-Based Online and Mixed-Mode Panels: Selected Recent Trends and Future Perspectives. *Social Science Computer Review*, 34(1), 3-7. https://doi.org/10.1177/0894439315579246

*<strong>Bottoni, G., & Fitzgerald, R.</strong> (2021). Establishing a baseline: Bringing innovation to the evaluation of cross-national probability-based online panels. *Survey Research Methods*, 15(2), 115–133. https://doi.org/10.18148/srm/2021.v15i2.7457

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

Bryda, G., & Costa, A. P. (2023). Qualitative research in digital era: Innovations, methodologies and collaborations. *Social Sciences*, 12(10), Article 570. https://doi.org/10.3390/socsci12100570

Büchi, M., Just, N., & Latzer, M. (2016). Modeling the second-level digital divide: A five-country study of social differences in Internet use. *New Media & Society*, 18(11), 2703–2722. https://doi.org/10.1177/1461444815604154

Callegaro, M., Baker, R., Bethlehem, J., Göritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (2014). Online panel research. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 1–22). John Wiley & Sons. https://doi.org/10.1002/9781118763520.ch1

Callegaro, M., Lozar Manfreda, K., & Vehovar, V. (2015a). Pre-fielding. In *Web survey methodology* (pp. 35–164). SAGE Publications.

Callegaro, M., Lozar Manfreda, K., & Vehovar, V. (2015b). Selected topics in web survey implementation. In *Web survey methodology* (pp. 191–230). SAGE Publications.

Callegaro, M., Lozar Manfreda, K., & Vehovar, V. (2015c). Survey research and web surveys. In *Web survey methodology* (pp. 3–34). SAGE Publications.

Centralna tehniška knjižnica Univerze v Ljubljani. (2025). DiKUL – Katalog informacijskih virov. https://viri.ctk.uni-lj.si/

*<strong>Chang, L., & Krosnick, J. A.</strong> (2009). National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly*, 73(4), 641–678. https://doi.org/10.1093/poq/nfp075

*<strong>Cho, S. K., LoCascio, S. P., Lee, K.-O., Jang, D.-H., & Lee, J. M.</strong> (2017). Testing the representativeness of a multimode survey in South Korea: Results from KAMOS. *Asian Journal for Public Opinion Research*, 4(2), 73–87. https://doi.org/10.15206/ajpor.2017.4.2.73

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10(1), 101–129. https://doi.org/10.2307/3001666

Cornesse, C., & Blom, A. G. (2023). Response quality in nonprobability and probability-based online panels. *Sociological Methods & Research*, 52(2), 879–908. https://doi.org/10.1177/0049124120914940

Cornesse, C., & Schaurer, I. (2021). The long-term impact of different offline population inclusion strategies in probability-based online panels: Evidence from the German Internet Panel and the GESIS Panel. *Social Science Computer Review*, 39(4), 687–704. https://doi.org/10.1177/0894439320984131

Cornesse, C., Felderer, B., Fikel, M., Krieger, U., & Blom, A. G. (2022a). Recruiting a probability-based online panel via postal mail: Experimental evidence. *Social Science Computer Review*, 40(5), 1259–1284. https://doi.org/10.1177/08944393211006059

*__Cornesse, C., Krieger, U., Sohnius, M.-L., Fikel, M., Friedel, S., Rettig, T., Wenz, A., Juhl, S., Lehrer, R., Möhring, K., Naumann, E., Reifenscheid, M., & Blom, A. G.__ (2022b). From German Internet Panel to Mannheim Corona Study: Adaptable probability-based online panel infrastructures during the pandemic. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 185(3), 773–797. https://doi.org/10.1111/rssa.12749

de Leeuw, E. D. (2008). Choosing the method of data collection. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 113–135). Lawrence Erlbaum Associates.

de Leeuw, E. D., & Nicholls, W. L., II. (1996). Technological innovations in data collection: Acceptance, data quality and costs. *Sociological Research Online*, 1(4), 23–37. https://doi.org/10.5153/sro.50

Daikeler, J., Bošnjak, M., & Lozar Manfreda, 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

*__Dever, J. A., Amaya, A., Srivastav, A., Roy, K., & Singleton, J. A.__ (2021). Fit for purpose in action: Design, implementation, and evaluation of the National Internet Flu Survey. *Journal of Survey Statistics and Methodology*, 9(3). https://doi.org/10.1093/jssam/smaa005

*__Dickie, M., Gerking, S., & Goffe, W. L.__ (2007). Valuation of non-market goods using computer-assisted surveys: A comparison of data quality from internet and RDD sample. https://cook.rfe.org/Survey_Comparison_3.pdf

Dillman, D. A. (2007). Introduction to tailored design. In *Mail and internet surveys: The tailored design method* (pp. 3–31). John Wiley & Sons.

DiSogra, C., & Callegaro, M. (2015). Metrics and design tool for building and evaluating probability-based online panels. *Social Science Computer Review*, 34(1), 26–40. https://doi.org/10.1177/0894439315573925

Eckman, S. (2015). Does the inclusion of non-Internet households in a web panel reduce coverage bias? *Social Science Computer Review*, 34(1), 41–58. https://doi.org/10.1177/0894439315572985

Eckman, S., Unangst, J., Dever, J. A., & Antoun, C. (2023). The precision of estimates of nonresponse bias in means. *Journal of Survey Statistics and Methodology*, 11(4), 758–783. https://doi.org/10.1093/jssam/smac019

Felderer, B., Repke, L., Weber, W., Schweisthal, J., & Bothmann, L. (2024). Predicting the validity and reliability of survey questions. *OSF Preprints*. https://doi.org/10.31219/osf.io/hkngd

Gaia, A., Sala, E., & Respi, C. (2025). Internet Coverage Bias in Web Surveys in Europe. *Survey Research Methods*, 19(2), 153–174. https://doi.org/10.18148/srm/2025.v19i2.8298

Goodman, A., Brown, M., Silverwood, R. J., Sakshaug, J. W., Calderwood, L., Williams, J., & Ploubidis, G. B. (2022). The impact of using the web in a mixed-mode follow-up of a longitudinal birth cohort study: Evidence from the National Child Development Study. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 185(3), 822–850. https://doi.org/10.1111/rssa.12786

*__Grönlund, K., & Strandberg, K.__ (2014). Online panels and validity: Representativeness and attrition in the Finnish eOpinion panel. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 86–103). John Wiley & Sons.

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

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). An introduction to survey methodology (Chapter 1). In *Survey methodology* (pp. 1–38). Wiley.

Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. (2021). Multilevel meta-analysis. In *Doing meta-analysis with R: A hands-on guide* (pp. 287–302). Chapman and Hall/CRC. https://doi.org/10.1201/9781003107347

*__Hemsworth, L. M., Rice, M., Hemsworth, P. H., & Coleman, G. J.__ (2021). Telephone survey versus panel survey samples assessing knowledge, attitudes and behavior regarding animal welfare in the red meat industry in Australia. *Frontiers in Psychology*, 12, Article 653620. https://doi.org/10.3389/fpsyg.2021.653620

*__Herman, P. M., Slaughter, M. E., Qureshi, N., & Lang, D. L.__ (2024). Comparing health survey data cost and quality between Amazon's Mechanical Turk and Ipsos' KnowledgePanel: An observational study. *Journal of Medical Internet Research*, 26(4), Article e56794. https://doi.org/10.2196/56794

Hernandez, K., & Faith, B. (2023). Online but still falling behind: Measuring barriers to internet use 'after access'. *Internet Policy Review*, 12(2), Article 1713. https://doi.org/10.14763/2023.2.1713

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

*__Høgestøl, A., & Skjervheim, Ø.__ (2014). *Norwegian citizen panel, 2013, first wave*. Methodology report. https://www.uib.no/sites/w3.uib.no/files/attachments/ncp-methodology-report-wave-1.pdf

*__Huggins, V., & Eyerman, J.__ (2001). Probability based Internet surveys: A synopsis of early methods and survey research results. In Research Conference for the Federal Committee on Statistical Methodology, Arlington, VA. https://nces.ed.gov/FCSM/pdf/2001FCSM_Huggins.pdf

Huijsmans, T., Harteveld, E., van der Brug, W., & Lancee, B. (2021). Are cities ever more cosmopolitan? Studying trends in urban-rural divergence of cultural attitudes. *Political Geography*, 86, 102353. https://doi.org/10.1016/j.polgeo.2021.102353

Idika, D. O., Owan, V. J., & Agama, V. U. (2023). The application of the nominal scale of measurement in research data analysis. *Prestige Journal of Education*, 6(1), 190–197.

Ji, L. J., Peng, K., & Nisbett, R. E. (2000). Culture, control, and perception of relationships in the environment. *Journal of Personality and Social Psychology*, 78(5), 943–955. https://doi.org/10.1037//0022-3514.78.5.943

*__Kaczmirek, L., Phillips, B., Pennay, D. W., Lavrakas, P. J., & Neiger, D.__ (2019). *Building a probability-based online panel: Life in Australia (CSRM Methods Series, No. 2/2019).* Centre for Social Research and Methods, Australian National University. https://csrm.cass.anu.edu.au/research/publications/building-probability-based-online-panel-life-australia

*__Kaufman, D. J., Baker, R., Milner, L. C., Devaney, S., & Hudson, K. L.__ (2016). A survey of U.S. adults' opinions about conduct of a nationwide Precision Medicine Initiative® cohort study of genes and environment. *PLOS ONE*, 11(8), e0160461. https://doi.org/10.1371/journal.pone.0160461

*__Kennedy, C., Mercer, A., Keeter, S., Hatley, N., McGeeney, K., & Gimenez, A.__ (2016). *Evaluating online nonprobability surveys: Vendor choice matters; widespread errors found for estimates based on Blacks and Hispanics*. Pew Research Center. https://www.pewresearch.org/methods/2016/05/02/evaluating-online-nonprobability-surveys/

Kocar, S., & Baffour, B. (2023). Comparing and improving the accuracy of nonprobability samples: Profiling Australian surveys. *Methods, Data, Analyses*, 17(2). https://doi.org/10.12758/MDA.2023.04

*__Kocar, S., & Biddle, N.__ (2023). Do we have to mix modes in probability-based online panel research to obtain more accurate results? *Methods, Data, Analyses*, 17(1), Article 11. https://doi.org/10.12758/mda.2022.11

Kocar, S., & Kaczmirek, L. (2023). A meta-analysis of worldwide recruitment rates in 23 probability-based online panels, between 2007 and 2019. *International Journal of Social Research Methodology*, 27(5), 589–604. https://doi.org/10.1080/13645579.2023.2242202

Kordzadeh, N., & Ghasemaghaei, M. (2021). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Lalla, M. (2017). Fundamental characteristics and statistical analysis of ordinal variables: A review. *Quality & Quantity*, 51(1), 435–458. https://doi.org/10.1007/s11135-016-0314-5

Lavrakas, P. J., Pennay, D., Neiger, D., & Phillips, B. (2022). Comparing probability-based surveys and nonprobability online panel surveys in Australia: A total survey error perspective. *Survey Research Methods*, 16(2), 241–266. https://doi.org/10.18148/srm/2022.v16i2.7907

*__Lee, S.__ (2006). An evaluation of nonresponse and coverage errors in a prerecruited probability web panel survey. *Social Science Computer Review*, 24(4), 460–475. https://doi.org/10.1177/0894439306288085

*__Leenheer, J., & Scherpenzeel, A. C.__ (2013). Does it pay off to include non-internet households in an internet panel? *International Journal of Internet Science*, 8(1), 17–29.

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79, 328–348. https://doi.org/10.1016/j.jesp.2018.08.009

*__Liu, S. T., Loomis, B. R., Kinsey, S. H., & Taylor, N. C.__ (2022). Development of a panel of US adult tobacco users to inform tobacco regulatory science. *Preventive Medicine Reports*, 28, 101827. https://doi.org/10.1016/j.pmedr.2022.101827

Lorenz, M., & Konečný, M. (2023). Digital archives as research infrastructure of the future. *Acta Informatica Pragensia*, 12(2), 327–341. https://doi.org/10.18267/j.aip.219

Lozar Manfreda, K., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web Surveys versus other Survey Modes: A Meta-Analysis Comparing Response Rates. *International Journal of Market Research*, 50(1), 79–104. https://doi.org/10.1177/147078530805000107

Lozar Manfreda, K., Couper, M., Vohar, M., Rivas, S., & Vehovar, V. (2002). Virtual selves and web surveys. In A. Ferligoj & A. Mrvar (Eds.), *Developments in social science methodology* (pp. 187–213). University of Ljubljana, FDV.

*__Lugtig, P., Das, M., & Scherpenzeel, A.__ (2014). Nonresponse and attrition in a probability-based online panel for the general population. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 135–153). John Wiley & Sons. https://doi.org/10.1002/9781118763520.ch6

*__MacInnis, B., Krosnick, J. A., Ho, A. S., & Cho, M.-J.__ (2018). The accuracy of measurements with probability and nonprobability survey samples: Replication and extension. *Public Opinion Quarterly*, 82(4), 707–744. https://doi.org/10.1093/poq/nfy038

Maslovskaya, O., & Lugtig, P. (2022). Representativeness in six waves of CROss-National Online Survey (CRONOS) panel. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 185(3), 851–871. https://doi.org/10.1111/rssa.12801

*__McMillen, R. C., Winickoff, J. P., Wilson, K., Tanski, S., & Klein, J. D.__ (2015). A dual-frame sampling methodology to address landline replacement in tobacco control research. *Tobacco Control*, 24(1), e10–e15. https://doi.org/10.1136/tobaccocontrol-2013-051436

*__Mercer, A., & Lau, A.__ (2023). *Comparing two types of online survey samples*. Pew Research Center. https://www.pewresearch.org/methods/2023/09/07/comparing-two-types-of-online-survey-samples/

Migliavaca, C. B., Stein, C., Colpani, V., Barker, T. H., Ziegelmann, P. K., Munn, Z., & Falavigna, M. (2022). Meta-analysis of prevalence: I^2 statistic and how to deal with heterogeneity. *Research Synthesis Methods*, 13(3), 363–367. https://doi.org/10.1002/jrsm.1547

Nayak, M. S. D. P., & Narayan, K. A. (2019). Strengths and weaknesses of online surveys. *IOSR Journal of Humanities and Social Sciences*, 24(5), 31–38.

Ormston, R., Martin, C., Rogers, L., Huskinson, T., Irvin, E., Rimmington, E., & Lynn, P. (2024). Financial and resource implications. In *Long term survey strategy: Mixed mode research report* (Chap. 10). Scottish Government, Chief Statistician, Digital Directorate. https://www.gov.scot/publications/mixed-mode-research-report-inform-scottish-government-long-term-survey-strategy/

Page, M. J., Higgins, J. P. T., & Sterne, J. A. C. (2024). Chapter 13: Assessing risk of bias due to missing results in a synthesis. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (Version 6.5, updated August 2024). Cochrane. https://www.training.cochrane.org/handbook

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. https://doi.org/10.1136/bmj.n71

Pang, J., & Capek, J. (2020). Factors influencing researcher cooperation in virtual academic communities based on principal component analysis. *Acta Informatica Pragensia*, 9(1), 4–17. https://doi.org/10.18267/j.aip.128

*__Pennay, D. W., Neiger, D., Lavrakas, P. J., & Borg, K.__ (2018). *The Online Panels Benchmarking Study: A total survey error comparison of findings from probability-based surveys and non-probability online panel surveys in Australia (CSRM Methods Series No. 2/2018).* Centre for Social Research and Methods.

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

Räsänen, P., Oksanen, A., Lehdonvirta, V., & Blank, G. (2024). Social media, web, and panel surveys: Using non-probability samples to study population characteristics. In *Advanced research methods for applied psychology: Design, analysis and reporting* (2nd ed., pp. 140–152). Routledge. https://doi.org/10.4324/9781003362715-13

Rasinski, K. A., Lee, L., & Krishnamurty, P. (2012). Question order effects. In H. Cooper (Ed.-in-Chief), P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology: Vol. 1. Foundations, planning, measures, and psychometrics* (pp. 229–248). American Psychological Association. https://doi.org/10.1037/13619-014

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 295–315). Russell Sage Foundation.

Rebenok, V., Rozhi, I., Petro, Y., Kozub, H., & Diachenko, N. (2024). Evolving information landscape: ICT's influence on societal digitalisation. *Multidisciplinary Science Journal*, 6, 2024ss0706. https://doi.org/10.31893/multiscience.2024ss0706

Rehm, J., Patra, J., Brennan, A., Buckley, C., Greenfield, T. K., Kerr, W. C., Manthey, J., Purshouse, R. C., Rovira, P., Shuper, P. A., & Shield, K. D. (2021). The role of alcohol use in the aetiology and progression of liver disease: A narrative review and a quantification. *Drug and Alcohol Review*, 40(5), 638–646. https://doi.org/10.1111/dar.13286

*__Revilla, M.__ (2013). Measurement invariance and quality of composite scores in a face-to-face and a web survey. *Survey Research Methods*, 7(1), 17–28. https://doi.org/10.18148/srm/2013.v7i1.5098

Ryan, K. L., Taylor, S. M., Lyle, J. M., Stark, K. E., & Tracey, S. R. (2024). On the Line and Online: Higher Non-Response to Web-Based Surveys Over-Represents Avid Recreational Fishers Compared With Telephone Surveys. *Fisheries Management and Ecology*, 32(3). https://doi.org/10.1111/fme.12752

Sakshaug, J. W., Wiśniowski, A., Perez Ruiz, D. A., & Blom, A. G. (2019). Supplementing small probability samples with nonprobability samples: A Bayesian approach. *Journal of Official Statistics*, 35(3), 653–681. https://doi.org/10.2478/JOS-2019-0027

Saris, W. E., & Gallhofer, I. N. (2007). Design, evaluation, and analysis of questionnaires for survey research. Wiley. https://doi.org/10.1002/9780470165195

Scherpenzeel, A. (2011). Data collection in a probability-based Internet panel: How the LISS panel was built and how it can be used. *Bulletin de Méthodologie Sociologique*, 109(1), 56–61. https://doi.org/10.1177/0759106310387713

*__Scherpenzeel, A. C., & Bethlehem, J. G.__ (2011). How representative are online panels? Problems of coverage and selection and possible solutions. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the Internet: Advances in applied methods and research strategies* (pp. 105–132). Routledge/Taylor & Francis Group.

*__Schonlau, M., van Soest, A., & Kapteyn, A.__ (2007). Are 'webographic' or attitudinal questions useful for adjusting estimates from web surveys using propensity scoring? *Survey Research Methods*, 1(3), 155–163. https://doi.org/10.18148/srm/2007.v1i3.70

Schwarz, N., Knäuper, B., & Oyserman, D. (2008). The psychology of asking questions. In E. de Leeuw, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 18–34). Taylor & Francis.

Selman, C. J., Lee, K. J., Whitehead, C. L., Manley, B. J., & Mahar, R. K. (2023). Statistical analyses of ordinal outcomes in randomised controlled trials: Protocol for a scoping review. *Trials*, 24, Article 72. https://doi.org/10.1186/s13063-023-07262-8

*__Seol, D.-H., Jang, D.-H., & LoCascio, S. P.__ (2023). RDD with follow-up texting: A new attempt to build a probability-based online panel in South Korea. *Asian Journal for Public Opinion Research*, 11(3), 257–273. https://doi.org/10.15206/ajpor.2023.11.3.257

*__Smith, T. W.__ (2003). An experimental comparison of Knowledge Networks and the GSS. *International Journal of Public Opinion Research*, 15(2), 167–179. https://doi.org/10.1093/ijpor/15.2.167

*__Smith, T. W., & Dennis, J. M.__ (2004). Comparing the Knowledge Networks web-enabled panel and the in-person 2002 General Social Survey: Experiments with mode, format, and question wordings (GSS Methodological Report No. 99). National Opinion Research Center, University of Chicago. https://gss.norc.org/content/dam/gss/get-documentation/pdf/reports/methodological-reports/MR99%20Comparing%20the%20Knowledge%20Networks%20Web-Enabled%20Panel%20and%20the%20In-Person%202002%20GSS.pdf

*Spijkerman, R., Knibbe, R., Knoops, K., Van de Mheen, D., & Van den Eijnden, R. (2009). The utility of online panel surveys versus computer-assisted interviews in obtaining substance-use prevalence estimates in the Netherlands. *Addiction*, 104(10), 1641–1645. https://doi.org/10.1111/j.1360-0443.2009.02642.x

Stadtmüller, S., Silber, H., Gummer, T., Sand, M., Zins, S., Beuthner, C., & Christmann, P. (2023). Evaluating an alternative frame for address-based sampling in Germany: The address database from Deutsche Post Direkt. *Methods, Data, Analyses*, 17(1), 29–46. https://doi.org/10.12758/mda.2022.06

Stahl, B. C., Timmermans, J., & Flick, C. (2017). Ethics of emerging information and communication technologies: On the implementation of responsible research and innovation. *Science and Public Policy*, 44(3), 369–381. https://doi.org/10.1093/scipol/scw069

*Stanley, M., Roycroft, J., Amaya, A., Dever, J. A., & Srivastav, A. (2020). The effectiveness of incentives on completion rates, data quality, and nonresponse bias in a probability-based internet panel survey. *Field Methods*, 32(2), 159–179. https://doi.org/10.1177/1525822x20901802

Struminskaya, B. (2014). Data quality in probability-based online panels: Nonresponse, attrition, and panel conditioning. Doctoral dissertation, Utrecht University. https://dspace.library.uu.nl/bitstream/1874/301751/3/struminskaya.pdf

*Struminskaya, B., de Leeuw, E., & Kaczmirek, L. (2016). Mode system effects in an online panel study: Comparing a probability-based online panel with two face-to-face reference surveys. *Methods, Data, Analyses*, 9(1), 3–56. https://doi.org/10.12758/mda.2015.001

Survey Quality Predictor. (2017). SQP coding instructions. Universitat Pompeu Fabra. http://sqp.upf.edu/media/files/sqp_coding_instructions.pdf

Symbaluk, D., & Hall, R. (2024). Surveys. In *Research methods: Exploring the social world in Canadian context* (Chap. 7). MacEwan University.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859–883. https://doi.org/10.1037/0033-2909.133.5.859

Tourangeau, R., Conrad, F. G., & Couper, M. P. (2013). Introduction. In *The science of web surveys* (pp. 1–10). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199747047.003.0001

*Unangst, J. J., Amaya, A. E., Sanders, H. L. P. S., Howard-Doering, J., Ferrell, A. R., Karon, S. L., & Dever, J. A. (2020). A process for decomposing total survey error in probability and nonprobability surveys: A case study comparing health statistics in US internet panels. *Journal of Survey Statistics and Methodology*, 8(1), 62–88. https://doi.org/10.1093/jssam/smz040

*Vaithianathan, R., Hool, B., Hurd, M. D., & Rohwedder, S. (2021). High-frequency internet survey of a probability sample of older Singaporeans: The Singapore Life Panel. *Singapore Economic Review*, 66(6), 1759–1778.

van der Schyff, K., Foster, G., Renaud, K., & Flowerday, S. (2023). Online privacy fatigue: A scoping review and research agenda. *Future Internet*, 15(5), Article 164. https://doi.org/10.3390/fi15050164

Vehovar, V., & Beullens, K. (2017). Cross-national issues in response rates. In D. L. Vannette & J. A. Krosnick (Eds.), *The Palgrave Handbook of Survey Research* (pp. 29–42). Springer International Publishing. https://doi.org/10.1007/978-3-319-54395-6_5

Vehovar, V., & Lozar Manfreda, K. (2017). Overview: Online surveys. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 143–161). SAGE.

Vehovar, V., Batagelj, Z., Lozar Manfreda, K., & Zaletel, M. (2002). Nonresponse in web surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 229–242). Wiley.

Vehovar, V., Smutny, Z., & Bartol, J. (2022). Evolution of social informatics: Publications, research, and educational activities. *The Information Society*, 38(5), 307–333. https://doi.org/10.1080/01972243.2022.2092570

Vehovar, V., Toepoel, V., & Steinmetz, S. (2016). Non-probability sampling. In C. Wolf, D. Joye, T. W. Smith, & Y.-C. Fu (Eds.), *The SAGE handbook of survey methodology* (pp. 329–345). SAGE.

Verhulst, B., & Neale, M. C. (2021). Best practices for binary and ordinal data analyses. *Behavior Genetics*, 51(3), 204–214. https://doi.org/10.1007/s10519-020-10031-x

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, 30(3), 261–293. https://doi.org/10.3102/10769986030003261

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

Vojíř, S., & Kučera, J. (2021). Towards re-decentralized future of the web: Privacy, security and technology development. *Acta Informatica Pragensia*, 10(3), 349–369. https://doi.org/10.18267/j.aip.169

Westland, H., Vervoort, S., Kars, M., & Jaarsma, T. (2025). Interviewing people on sensitive topics: Challenges and strategies. *European Journal of Cardiovascular Nursing*, 24(3), 488–493. https://doi.org/10.1093/eurjcn/zvae128

Williams, M. S., Ebel, E. D., & Wagner, B. A. (2007). Monte Carlo approaches for determining power and sample size in low-prevalence applications. *Preventive Veterinary Medicine*, 82(1-2), 151–158. https://doi.org/10.1016/j.prevetmed.2007.05.015

*Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly*, 75(4), 709–747. https://doi.org/10.1093/poq/nfr020

Acta Informatica Pragensia is published by the Prague University of Economics and Business, Czech Republic | eISSN: 1805-4951