Probability-Based Web Panels for Official Statistics: Basic Insights and Analysis of the Bias of Survey Estimates

Gregor Čehovin[1] | *University of Ljubljana, Ljubljana, Slovenia*
Vasja Vehovar[2] | *University of Ljubljana, Ljubljana, Slovenia*

Received 8.7.2025 (revision received 24.9.2025), **Accepted** (reviewed) 28.10.2025, **Published** 12.6.2026

Abstract

The past three decades of survey research have revealed a shift from traditional probability-based surveys to various types of web surveys. In official statistics, the probability-based web panels (PWPs), for which respondents are recruited once and then incentivized for repeated participation in web surveys, play a particularly important role in this process. This article provides insights into the use of PWPs for official statistics. First, a survey of European Union national statistics offices revealed that one-quarter had already implemented or were planning to implement PWPs; the main barriers to their implementation were lack of knowledge and expertise. In addition, we evaluated the quality of the estimates in the Slovenian PWP (1KA Panel) that replicated questions from 12 traditional probability-based surveys. The findings showed that 205 of 651 PWP estimates (31%) exhibited relative bias exceeding 10%. Biases varied substantially across survey topics, indicating the selective suitability of PWPs for official statistics.

Keywords	***DOI***	***JEL code***
Probability-based web panels, survey estimates, relative bias, nonresponse error, measurement error, coverage error, processing error	*https://doi.org/10.54694/stat.2025.28*	*C83, Y80*

INTRODUCTION

Since the 1970s, the evolution of digital technologies has steadily transformed the survey research process from *traditional survey modes* – i.e., face-to-face, telephone, and mail surveys – to web surveys. Survey science (Groves et al., 2009), however, was originally grounded on these traditional modes,

[1] Centre for Social Informatics, Faculty of Social Sciences, University of Ljubljana, Kardeljeva ploščad 5, 1000 Ljubljana, Slovenia. Corresponding author: e-mail: gregor.cehovin@fdv.uni-lj.si. ORCID: <https://orcid.org/0000-0002-9431-3694>
[2] Centre for Social Informatics, Faculty of Social Sciences, University of Ljubljana, Kardeljeva ploščad 5, 1000 Ljubljana, Slovenia. E-mail: vasja.vehovar@fdv.uni-lj.si. ORCID: <https://orcid.org/0000-0002-3253-7959>.

typically implemented using the design-based sampling theory (e.g., Cochran, 1977; Särndal et al., 1992). We hereafter refer to them as *traditional probability-based surveys* (TPSs), a term that also includes contemporary TPSs that combine traditional modes with web surveys (i.e., mixed mode).

In recent years, rising TPS costs and declining response rates have been addressed through inexpensive, flexible web surveys (Callegaro et al., 2015), especially nonprobability ones (Callegaro et al., 2014; Golini and Righi, 2024; Vehovar et al., 2016). *Probability-based web panels* (PWPs) have emerged as viable alternatives for official statistics grounded in probability principles. Respondents are recruited via probabilistic sampling and then invited to participate in recurring web surveys (e.g., monthly for two years), typically with incentives (Couper, 2017; Hays et al., 2015). PWPs reduce per-survey costs by requiring only one recruitment drive, enable faster data collection, and maintain probability principles.

Although PWPs involve lower costs and facilitate more flexible data collection, they generally yield lower response rates than TPSs (e.g., Kocar and Kaczmirek, 2024). This increases the risk of nonresponse bias – the difference between a survey estimate and its true (population) value caused by nonresponse. The issue is critical for official statistics, where unbiased estimates are essential. Whether the operational advantages of PWPs offset potential declines in estimate quality remains unclear. Empirical evidence concerning the role of PWPs in official surveys, particularly the effects of transitioning from TPSs to PWPs, is limited. To address these gaps, we pose the following research questions (RQs):

- RQ1: What are the practices, plans, and perceptions of PWPs for official statistics?
- RQ2: What is the extent of bias of the estimates derived from PWPs compared to those obtained from TPSs?

To address these questions, we surveyed *national statistical institutes* (NSIs) in the European Union (EU) for RQ1 and conducted a comprehensive study of Slovenian official statistics for RQ2 by comparing estimates from 12 TPSs with those from a parallel PWP.

1 PROBABILITY-BASED WEB PANELS

PWPs possess three essential characteristics: (a) long-term respondent participation across multiple studies, based on informed consent and incentives; (b) primary data collection via the web, supplemented by traditional modes for internet nonusers but not for recontacting web survey nonrespondents; and (c) probability-based sampling, requiring known, positive inclusion probabilities for all population units and a list of all units in the target population or a surrogate sample, such as an area sample. Unfortunately, constructing and exploiting such lists is resource-intensive. In general population surveys, the absence of email address lists necessitates traditional sampling methods, including population registers, address databases, telephone directories, or random digit dialing, typically via postal, telephone, or in-person approaches. Recruitment may also use the existing TPSs ("piggybacking"). PWPs thus combine the rigor of probability-based sampling with the cost-efficiency of web surveys.

Several PWPs have been established in EU countries, including the Longitudinal Internet Studies for the Social Sciences (LISS) panel in the Netherlands in 2007, followed by initiatives such as the German Internet Panel (GIP), GESIS Panel, Étude Longitudinale par Internet Pour les Sciences Sociales (ELIPSS), and Norwegian Citizen Panel (NCP) (Bottoni and Sommer, 2019; Kocar and Kaczmirek, 2024). The European Social Survey (ESS) has also conducted three pilot PWP studies (the CROss-National Online Survey [CRONOS] projects) and plans to integrate them into ESS surveys (European Social Survey, 2025).

In the United States (U.S.; Couper, 2017), PWPs began in the commercial sector in 1998 with KnowledgePanel (formerly KnowledgeNetworks), followed by academic and nonprofit initiatives such as the Face-to-Face Recruited Internet Survey Platform (FFRISP), NORC's AmeriSpeak, Pew Research Center's American Trends Panel, RAND Corporation's American Life Panel, and the Understanding America Study (UAS). Other commercial panels include the SSRS Opinion Panel and Gallup Panel.

The U.S. Census Bureau (2025) also established the Household Trends and Outlook Pulse Survey (HTOPS) for official statistics. Several PWPs have been established outside the EU and U.S. (Kocar and Kaczmirek, 2024).

PWPs differ in recruitment, operations, incentives, eligible respondent ages, and size (e.g., Blom et al., 2017; Bosnjak et al., 2018). Most comprise a few thousand participants and rarely exceed 10 000. To improve coverage, some PWPs, such as CRONOS (European Social Survey, 2023) and LISS (Centerdata, 2024), provide participation options for individuals without internet access through in-person, postal, or telephone surveys; however, the overall cost and time implications remain uncertain (e.g., Bach et al., 2024; Kocar and Biddle, 2023).

Comparisons of PWPs and TPSs are methodologically complex, as mode effects and other measurement errors must be distinguished from sampling, noncoverage, processing, and nonresponse errors (Berrens et al., 2003; Cornesse et al., 2020; Struminskaya et al., 2015). Thus, it is not surprising that findings are often inconclusive. Some studies show that PWPs approximate TPS accuracy, with measurement instruments performing similarly or better for sensitive questions, though concerns persist about nonresponse bias and coverage limitations (Cornesse et al., 2020; Pennay et al., 2018; Yeager et al., 2011). Similarly, Bosnjak et al. (2018) revealed minimal differences between the GESIS Panel and related TPSs. However, some studies identified notable discrepancies between PWPs and TPSs (e.g., Ivanovska et al., 2025; Mercer and Lau, 2023; Pforr and Dannwolf, 2017; Struminskaya et al., 2015).

These methodological challenges have limited the adoption of PWPs for official statistics, and caution has generally prevailed. Empirical applications remain rare, primarily concentrated in health statistics (e.g., Ivanovska et al., 2025; Lemcke et al., 2024; Peytchev, 2025), with additional studies on sociodemographic indicators (e.g., Seol et al., 2023), consumer expenditure (Graf et al., 2025), and civic, economic, and well-being measures (Kocar and Biddle, 2023). Many contributions, however, remain conceptual (Bethlehem, 2014; Svensson, 2014). Nevertheless, the suspension of in-person surveys during COVID-19 renewed interest in PWPs as an alternative to TPSs. Recent methodological innovations further support their implementation, including knock-to-nudge (K2N), in which interviewers visit households to encourage participation (Smith, 2022), and electronic questionnaire delivery (EQD), which provides customized tablets to simplify online survey completion (Fitzgerald, 2022).

1.1 Response rates

PWPs face significant challenges, including low response rates, mode effects, panel conditioning, and attrition, all of which may compromise data quality (Kennedy et al., 2016). Response rates – the proportion of individuals invited who participate (American Association for Public Opinion Research, 2023) – and corresponding nonresponse rates are a key concern. Cumulative response rates (i.e., *overall response rates* [ORRs]), which account for all recruitment stages, are a central PWP nonresponse indicator (Callegaro and DiSogra, 2008).

Historically, TPS response rates of 70–80% were considered benchmarks by institutions such as the ESS (Koch et al., 2010), the Organisation for Economic Co-operation and Development (OECD, 2014), and the United States Office of Management and Budget (OMB, 2016). Although these benchmarks have been largely abandoned (Vehovar and Beullens, 2018), many TPS, particularly for official statistics, still exceed 50% (Jabkowski and Kołczyńska, 2020; Vehovar and Beullens, 2018), though this is increasingly rare in developed countries.

By contrast, PWP response rates are typically less than half the rate observed in comparable face-to-face surveys. In Slovenia, face-to-face recruitment yields about 50%, but only half of these respondents join the corresponding PWP. The Slovenian CRONOS-1 pilot panel recorded a 54% initial response rate linked to the 2018 ESS face-to-face survey, yet only 43% participated in the PWP, resulting in an ORR of 23% (Bottoni and Sommer, 2019). A similar pattern appeared in CRONOS-2 (Bottoni, 2023; Maslovskaya and Lugtig, 2022). Postal recruitment without supplemental sampling of internet nonusers generally yields ORRs below 20%, as in the CRONOS surveys (1KA Panel, 2025).

A study of 23 PWPs in 15 countries found that all but four reported ORRs below 20% (Kocar and Kaczmirek, 2024). In the early years of PWP development, panels such as the LISS and FFRISP panels, which used in-person recruitment and support for internet nonusers, achieved ORRs near 40% (Sakshaug et al., 2009; Scherpenzeel and Bethlehem, 2011), though such levels are now unlikely in developed countries. Commercial PWPs typically report even lower, often single-digit, response rates (Kennedy et al., 2016; Mercer et al., 2018; Olson et al., 2021; Pasek, 2016). Persistently low response rates therefore remain a key barrier to PWP use in official statistics.

1.2 Bias of estimates

Within the Total Survey Error (TSE) framework, survey error is classified into sampling and non-sampling components (Groves et al., 2009). Sampling error arises from estimating population parameters based on a sample and is typically quantified using variance measures. Non-sampling error refers to various deviations from true population values, with nonresponse and measurement errors typically most notable, along with coverage and processing errors. The present study does not estimate these error components separately, but their combined effect.

Formally, the bias of an estimate can be formulated as follows:

$$\mathrm{Bias}(\bar{y}) = \mathrm{E}\left[\bar{y}\right] - \bar{Y}, \tag{1}$$

where $\bar{y}$ is the estimate, $E[\bar{y}]$ is the expected value of this estimate, and $\bar{Y}$ is the true value (Groves et al., 2009). Typically, the *absolute value* of this bias is applied, which also holds for the corresponding estimate of the bias($\bar{y}$), i.e.,

$$\mathrm{bias}(\bar{y}) = \bar{y} - \bar{Y}. \tag{2}$$

For example, a Pew Research Center analysis comparing 28 demographic and lifestyle benchmarks across three U.S. PWPs revealed that the estimates of absolute bias ranged from 2.3 to 3.0 percentage points, with an overall average of 2.6 points (Mercer and Lau, 2023). An Australian PWP study reported average absolute bias of 3.6 points for substantive and 5.9 for demographic items (Pennay et al., 2018).

Comparative studies often assess *relative bias* (*RB*), typically expressed as a proportion. When *RB* is estimated from a single survey, we denote this estimate as *RB*' (Groves et al., 2009):

$$\mathrm{RB}' = \frac{\mathrm{bias}(\bar{y})}{\bar{Y}}. \tag{3}$$

RB' is defined only when $\bar{Y} > 0$. The estimate $\bar{y}$ typically derives from the PWP, while the reference value $\bar{Y}$ in the numerator is taken from an external benchmark (e.g., a census or administrative source) or its estimate, such as a high-quality estimate from TPS, which we denote $\bar{Y}'$. Unlike absolute bias, *RB* reflects the size effect: an identical absolute bias of 2 percentage points corresponds to an *RB* of 40% when the population share is 5% (estimate 7%) but only 4% when the share is 50% (estimate 52%).

A meta-analysis of 137 health-related survey items from 14 studies reported a median *RB*' of 12.7% for PWP estimates, with *RB*' ranging from 2.2% for doctor treatment to 23.6% for disability-related estimates (Ivanovska et al., 2025). An analysis of six waves of the Life in Australia PWP across 18 benchmarked items (e.g., health status, home ownership, and life satisfaction) reported median *RB*' values ranging from 4.7% to 6.4% (Kocar and Biddle, 2023). Recalculation of *RB*' for the above-mentioned Pew Research Center study (Mercer and Lau, 2023) showed a median *RB*' of 13.0% and a range of 0%–800% (Ivanovska et al., 2025).

Researchers often overemphasize response rates while underrepresenting other quality indicators such as nonresponse bias, sampling variance, response quality, and costs. Relationships among these indicators are complex, particularly between nonresponse rates and bias (Groves and Peytcheva, 2008), and further complicated by cost considerations (Callegaro et al., 2015; Vehovar and Beullens, 2018). In this context, a consensus statement by survey methodologists (Maslovskaya et al., 2025) recommends prioritizing sample composition and representativeness over raw response rates. Nonetheless, response rates remain the most visible quality parameter when the credibility of national statistics is challenged, as in the United Kingdom Labor Force Survey (LFS), for which response rates are critically low (Casey, 2024; Office for National Statistics, 2025).

2 PROBABILITY-BASED WEB PANELS FOR OFFICIAL STATISTICS

2.1 Methods

To assess PWP use in official statistics, we surveyed NSIs in EU member states and those engaged with Eurostat processes. The survey, sponsored by the University of Ljubljana (UL) and Statistical Office of the Republic of Slovenia (SORS), was conducted from October 11, 2022, to January 25, 2023. SORS provided a contact list. Invitations were sent to methodology section heads in 31 NSIs (27 EU states plus Iceland, Kosovo, Switzerland, and Norway). Reminders followed in October and December. Anonymity was ensured by omitting identifying information and internet protocol addresses.

Nineteen NSIs responded (61% response rate). The questionnaire addressed current practices, plans for PWPs, and perceptions of PWP-related factors. The results offer descriptive insights into the PWP landscape in the EU. Further technical details are available in Vehovar et al. (2023).

2.2 Results

Of the 19 NSIs, five (26%) reported engagement with PWPs. Two had implemented PWPs: one since 2000 using an internal panel recruited face-to-face or by telephone for 12 household and person surveys (response rate 6%–10%, no incentives, personal interviews for non-internet users); and one since 2001 using an external panel of 5 000–10 000 participants recruited by post or face-to-face, with devices and internet access provided to non-internet users (response rate 11%–20%, with incentives). Three NSIs were developing or considering PWPs: one in conceptual development targeting 10 000–20 000 members with a 21%–30% response rate, and two exploring implementation, including one planning an internal PWP.

Table 1 Perceived barriers to PWP adoption among NSIs not yet engaged in PWP implementation (n = 11; 1 = strongly disagree to 7 = strongly agree)

Statement (perceived barriers)	Mean	Standard deviation
Lack of knowledge and expertise regarding this topic	5.6	1.03
Low sample representativeness (e.g., noncoverage)	5.1	1.64
Too little evidence in the literature that this works for official statistics	5.0	1.41
Low response quality (e.g., validity and reliability)	4.6	1.44
Potential surveys too specific to be included in such panels	4.4	1.50
Low response rates	3.9	1.38
Managerial complexity	3.8	1.40
Relatively little cost savings	3.6	1.04

Source: Original data and analysis by the authors

For the above five NSIs that had engaged with PWPs, perceived advantages and disadvantages compared with TPSs were assessed on a 7-point scale (1 = strongly disagree, 7 = strongly agree); the values reported below are means. Advantages included ease of repeated observations (6.0), simplified data processing (5.8), faster fieldwork (5.6), simplified management (5.6), lower costs (5.4), easier sampling (5.2), and better response quality (5.2). Disadvantages were panel conditioning (5.4), panel attrition (5.4), lower representativeness (4.8), coverage problems for non-internet users (4.8), nonresponse error (4.6), lower response quality (4.6), survey topics being too specific (4.2), and lower overall response rates (4.0).

Among the 14 NSIs (74%) not engaged with PWPs, 11 assessed barriers to adoption, also on the same 7-point scale (Table 1). The most cited were lack of knowledge and expertise (5.6), low representativeness (5.1), insufficient evidence of suitability for official statistics (5.0), and low response quality (4.6).

3 BIAS IN PROBABILITY-BASED WEB SURVEYS

To address potential bias in the PWP estimates, we conducted a comprehensive study using estimates from TPSs in the context of Slovenian official statistics.

3.1 Sample and recruitment strategies

We analyzed data collected from October 2022 to July 2023, during which a Slovenian PWP (i.e., the 1KA Panel,[3] run by UL's Faculty of Social Sciences) operated concurrently with multiple TPSs. Data were collected across five waves of the PWP-0 (December 5, 2022–February 17, 2023), 1 (February 10, 2023–April 5, 2023), 2 (April 3, 2023–May 14, 2023), 3 (May 19, 2023–July 3, 2023), and 4 (June 14, 2023–July 23, 2023) – as well as during a parallel recruitment Wave 0' (October 26, 2022–January 23, 2023) to facilitate estimate comparisons. Notably, no surveys pertinent to this analysis were conducted during Wave 3, which focused exclusively on topics outside the scope of official statistics.

The panel included 1 628 Slovenian residents aged 18 or older, recruited via two TPS surveys (i.e., piggybacking) – the Comparative Study of Electoral Systems (CSES) and the Slovene Public Opinion Survey (SJM) 2022/2 – as well as through fresh recruitment from the Central Population Register. All invitations for the fresh sample were sent by post; no web pages or email addresses were used for frame construction or contact. All recruitment sources drew exclusively from population registers. A random systematic sample with implicit stratification by 12 statistical regions and five settlement types ensured equal selection probabilities across all three samples. The fresh sample contributed 717 panelists (20%) from a 3 600-unit sample. Piggybacking added 413 panelists from a 3 000-unit CSES sample and 498 from a 2 500-unit SJM 2022/2 sample. Wave 1 included all 1 628 consenting panelists. Waves 2–4 targeted the same cohort, adjusted for 15 opt-outs.

All waves employed push-to-web methods, with initial wave recruitment via postal invitations offering conditional €10 gift cards for Waves 0 and 0' and unconditional €5 cards for Waves 1–4; email was used solely for reminders. Wave 0 used two postal reminders and a conditional €10 gift card for 2 800 units, 1 540 of which were assigned to incentive experiments. Wave 0' used three postal reminders and a conditional €10 gift card for 2 500 units. Waves 1–4 offered unconditional €5 gift cards to all panelists, including prior nonrespondents. These methods yielded an 18% overall recruitment rate, with register-based recruitment alone achieving 20% and per-wave response rates ranging from 87% to 93%. Wave 1 achieved a 17% ORR with 1% breakoff; across Waves 1–4, the ORR was 16%. In Wave 4, 1,439 respondents participated, representing 88% of recruited panelists – within the 70–90% retention typical of high-quality panels – and 16% of the initial gross sample. Attrition remained low: only 1% of the initial sample (6% of panelists) dropped out between Waves 1 and 4. Completion rates were high: 81% completed all four waves, 9% three, 4% two, 4% one, and 3% none.

[3] More information about the 1KA Panel is available at: <*https://panel.1ka.si/?lang_id=2*>.

We compared estimates from various TPSs with PWP Waves 0–4 (i.e., the 1KA Panel) conducted from October 2022 to July 2023. The PWP relied exclusively on web surveys, whereas the TPSs typically employed mixed-mode surveys, combining web with face-to-face, telephone, or mail surveys. In TPSs, the second mode was applied only as a follow-up for nonrespondents. Each TPS drew its initial sample from a single frame based on population registers; no web pages or email addresses were used for frame construction, so each unit appeared only once in the sample. We compared 12 survey sets: 9 for SORS and 3 for the National Institute of Public Health (NIPH) (see Table 2). SORS and NIPH decided to include only selected estimates from these TPSs in the PWP survey; the number of variables selected appears in the final column of Table 2. The corresponding PWP questionnaires were developed using the 1KA[4] software.

Table 2 Overview of the 12 survey sets implemented for the TPSs and replicated in the PWP (the 1KA Panel)

Survey set acronym: full survey set name	TPS provider	TPS mode*	PWP wave	Variable count
Consumers: Consumer opinion	SORS	web, CATI	0′	8
Less salt: Monitoring excreted sodium, potassium, and iodine	NIPH	CAPI	0′	26
COVID: Work and living conditions	SORS	web, CATI	0	58
Adult learning: Adult learning and education	SORS	web, CAPI	0	143
SILC: Living conditions survey (persons)	SORS	CAPI, CATI	1	57
SILC (HH): Living conditions survey (households)	SORS	CAPI, CATI	1	54
Tourism: Tourist trips – domestic population	SORS	web, CATI	2	24
Tourism (HH): Tourist trips – domestic population (households)	SORS	web, CATI	2	16
Drugs: National survey on tobacco, alcohol, and other drugs	NIPH	web, CAPI	2	141
ICT: ICT use in households and by individuals	SORS	web, CAPI	2	39
Activity: Active and inactive population (LFS)	SORS	CAPI, CATI	4	17
Mental health: National survey on attitudes toward mental health	NIPH	web, PAPI	4	68

Note: * the second TPS mode was used only for nonrespondents; no additional sampling frames were introduced. SORS = Statistical Office of the Republic of Slovenia, NIPH = National Institute of Public Health of Slovenia, CAPI = Computer-Assisted Personal Interviewing, CATI = Computer-Assisted Telephone Interviewing, PAPI = Pen-and-Paper Personal Interview, ICT = Information and Communications Technology.
Source: Original data and analysis by the authors

3.2 Bias calculations

In this study, TPS estimates were used as the benchmark under the assumption that they are unbiased, as they represent the best available TPSs in Slovenia. We note that this approach has limitations, as TPS estimates may include some error and may also reflect mode or measurement effects. For the bias analysis, we evaluated the bias for all PWP estimates using Formula (2), assuming $\bar{Y}'$ from TPS as an unbiased estimate of the population value $\bar{Y}$. We then calculated RB' with Formula (3), expressing RB' values as percentages, and considered bias($\bar{y}$) statistically significant at $p < .05$. Analyses were conducted with weighted data in IBM SPSS Statistics and R (R Core Team, 2024). Significance was tested with Welch's t-test (i.e., comparing means from two independent samples), and standard errors were calculated using the Taylor linearization method for the ratio estimator, as weights were applied.

[4] *<https://www.1ka.si/d/en>.*

We used the Survey Weighting GUI application (Štrlekar and Vehovar, 2025) for weighting via the raking method, implemented with R's *anesrake* package (Pasek, 2022). Auxiliary variables included gender, age, region, settlement type, and education, with population distributions obtained from SORS. We trimmed weights above 5 to stabilize standard errors, normalized them to a mean of 1 and applied identical weighting to both PWP and TPS data.

3.3 Results

Throughout the remainder of this paper, we only interpret *RB*' that were related to statistically significant biases ($p < .05$). Table 3 shows, for 12 survey sets, the proportion of variables with *RB*' exceeding thresholds of 5%, 10%, and 20%. For subsequent analyses, we focused on *RB*' > 10% to identify substantial bias variation across survey sets. The proportion of variables exceeding this threshold ranged from 0% to 69%. Overall, 205 of 651 variables (31%) had *RB*' > 10%. The median *RB*' across all 651 variables was 16%.

Table 3 The 12 survey sets showing variable counts, median RB', and proportions of variables with RB' with corresponding biases statistically significant at p < .05 and exceeding thresholds (5%, 10%, 20%), listed in descending order of the proportion with RB' > 10%.

Survey set	Variable count	Median *RB'*	*RB'* > 5%		*RB'* > 10%		*RB'* > 20%	
		Proportion	Count	Proportion	Count	Proportion	Count	Proportion
ICT	39	22%	29	74%	27	69%	20	51%
SILC	57	23%	43	75%	35	61%	28	49%
SILC (HH)	54	14%	35	65%	31	57%	19	35%
Activity	17	39%	11	65%	9	53%	7	41%
Tourism (HH)	16	29%	7	44%	7	44%	6	38%
Less salt	26	21%	13	50%	10	38%	8	31%
Adult learning	143	14%	35	24%	33	23%	24	17%
Drugs	141	20%	33	23%	29	21%	21	15%
Mental health	68	10%	15	22%	14	21%	9	13%
COVID	58	11%	10	17%	8	14%	6	10%
Tourism	24	7%	2	8%	2	8%	1	4%
Consumers	8	4%	4	50%	0	0%	0	0%
All survey sets	651	16%	237	36%	205	31%	149	23%

Note: Refer to Table 2 for full names and providers of the survey sets.
Source: Original data and analysis by the authors

Variation in the proportion of estimates with *RB*' exceeding 10%, as well as in median *RB*' values within survey sets, was primarily associated with the survey topic. One survey set contained no such estimates (Consumers), while Tourism (individuals) and COVID constituted only 8% and 14% of the variables where *RB*' exceeded 10%. In 9 out of 12 sets, more than 20% of the estimates had *RB*' > 10%. For survey sets, more than 50% of estimates exceeded this threshold: Activity (53%), SILC (households 57%, persons 61%), and ICT (69%).

Variation in the proportion of estimates with *RB'* values exceeding 10% could be attributed to nonresponse bias, but also to other sources of errors (e.g., noncoverage, processing error, measurement error). Estimates from the PWPs were most comparable to the TPS surveys conducted using push-to-web approaches (see Table 2), with *RB'* > 10% for Tourism (8% of the estimates), COVID (14%), Drugs (21%), Mental health (21%), and Adult learning (23%) (Table 3). However, the ICT survey set, using a web survey plus Computer-Assisted Personal Interviewing (CAPI) follow-ups for nonrespondents in the TPS implementation, had the highest proportion of variables with *RB'* > 10% (69%). High *RB'* values were also found in TPS survey sets administered in person, particularly for SILC (61%), SILC (HH) (57%), Activity (53%), and Less salt (38%), where in-person administration may have amplified response differences, mainly via higher response rates.

A summary comparison of the estimates indicated that relative to TPS respondents, PWP respondents represented a distinct demographic—potentially younger and more digitally engaged – with a higher proportion residing in urban areas. These differences may have also contributed to the observed biases. Additionally, PWP respondents, on average, reported lower levels of satisfaction with life, relationships, and social trust. Their health-related differences included reduced health awareness, lower fruit and vegetable intake, poorer self-perceived health, longer-term medical conditions, but fewer disabilities. They had higher internet usage, higher employment, and lower retirement rates. PWP respondents also showed lower home ownership, more rental accommodation, and more single-person households. Educational attainment and interest in further learning, especially informal or online learning, were higher. Financial hardship was more common and was characterized by material deprivation and delayed payments. Travel was more frequent, longer, and more expensive.

4 DISCUSSION

In this section, we provide a summary evaluation of the RQs and highlight the limitations of the study and future research opportunities.

4.1 The use of probability-based panels for official statistics

In relation to RQ1, the survey of NSIs in the EU indicated that the web survey mode was only partially used for traditional longitudinal panels, demonstrating that comprehensive adoption across household and person surveys continues to present a significant challenge. In this context, PWPs constitute an even more advanced application of web-based data collection. Although adoption remains limited, approximately one-quarter of NSIs reported that they had implemented or initiated the implementation of PWPs. Nonetheless, substantial limitations and perceived barriers persist, as outlined in Section 2.2. The five NSIs that commenced the PWP adoption identified several operational advantages (e.g., ease of repeated observations, simplified survey procedures, increased speed), while the primary disadvantages were directly or indirectly associated with the risks of introducing bias into estimates due to panel conditioning, attrition, compromised representativeness, noncoverage, nonresponse, and reduced response quality. The NSIs that had not yet adopted PWPs expressed caution due to similar concerns; however, the most frequently cited impediments were a lack of technical expertise and insufficient scientific literature addressing these issues.

In addition, both adopters and non-adopters of PWPs recognized that certain official surveys were unsuitable for the PWP integration. Although this was ranked among the less serious concerns, it remains important for specific surveys and necessitates thorough methodological consideration. Often, for example, PWPs have sample sizes that are too small, or complex data collection is required (e.g., biological samples; Couper, 2017).

The issues identified above partially reflect the general findings in the literature on PWPs (Cornesse et al., 2020; Mercer et al., 2018). In particular, repeated participation can lead to panel conditioning, altering respondent behavior, attrition, and undermining longitudinal consistency (Kennedy et al., 2016; Struminskaya and Bosnjak, 2021). Low response rates heighten the risk of nonresponse bias; however, response rate alone is a weak quality indicator, necessitating alternative assessments (e.g., benchmark comparisons, R-indicators). Very low response rates may reduce confidence in survey results unless accompanied by transparent evidence of representativeness. For the PWP or web-first approaches in official statistics, providing nonresponse-bias diagnostics is recommended to support stakeholders' confidence in the results (Cornesse and Bosnjak, 2018; Maslovskaya et al., 2025).

The appropriateness of PWPs varies by topic and mode, requiring methodological refinements to reduce bias without escalating costs (Bethlehem, 2014; Callegaro et al., 2015; Cornesse et al., 2020). However, limited expertise in advanced statistical methods and a scarcity of robust, peer-reviewed research on mitigating self-selection and noncoverage biases have hindered their wider adoption by NSIs (Svensson, 2014). Although PWPs are obviously less suitable when they involve high nonresponse or coverage errors, emerging approaches, such as adaptive survey designs and dual-frame methodologies, offer promising avenues for balancing costs and errors, emphasizing the need for comprehensive optimization beyond merely improving response rates or reducing bias (Schouten et al., 2017; Tourangeau et al., 2017; Vannieuwenhuyze, 2014).

As described in Section 2, PWPs have been implemented in multiple countries, although their integration into official statistics remains limited and uneven. A recent meta-analysis (Kocar and Kaczmirek, 2024) documents global adoption patterns and challenges. From a broader perspective, PWPs are unlikely to fully replace TPSs in the foreseeable future. Instead, official statistics increasingly adopt web-first mixed-mode designs, in which web surveys are supplemented by targeted interviewer follow-ups (e.g., CATI or CAPI) among nonrespondents or offline populations. This strategy preserves most cost and speed advantages of web surveys while reducing coverage and nonresponse issues, leaving TPSs essential for topics in which PWPs alone remain inadequate (Vannieuwenhuyze, 2014).

4.2 Biases of the estimates in probability-based web surveys

With respect to RQ2, concerning the extent of the bias in estimates derived from PWPs compared to those from TPSs, we conducted a comprehensive empirical study of Slovenian official statistics, and the results indicated a mixed pattern.

Several TPS survey sets exhibited a high proportion of estimates with *RB'* exceeding 10%, particularly for ICT (69%), SILC (61%), SILC (HH) (57%), and Activity (53%). These findings suggest that surveys involving complex socioeconomic or behavioral indicators are more susceptible to bias. This observation is consistent with Cornesse et al. (2020), who reported that PWPs tend to yield estimates comparable to TPSs for general attitudinal measures but display greater bias for behavioral or domain-specific variables, such as the ICT usage and socioeconomic indicators. Notably, the ICT survey demonstrated the highest proportion of estimates with *RB'* > 10% (69%), primarily attributable to the exclusion of internet nonusers from PWPs, as well as to the overrepresentation of individuals with higher levels of technological access or proficiency. The bias in ICT estimates thus reflects pronounced nonresponse and undercoverage issues (e.g., Groves and Peytcheva, 2008), as well as selection bias (e.g., Scherpenzeel and Bethlehem, 2011). Similar discrepancies were observed by Šoštarič (2020) for the CRONOS-1 panel, for which variables related to internet use exhibited high deviations compared to TPS estimates.

In contrast, certain survey sets showed comparatively low proportions of *RB'* > 10%, particularly for Consumers (0%), Tourism (8%), and COVID (14%). This pattern indicates that surveys involving less sensitive or cognitively simpler topics (e.g., consumer opinions) are less affected by PWP specifics. Conversely, surveys addressing sensitive domains (e.g., health, wealth, or life satisfaction) tend to encounter biases, warranting careful methodological consideration. However, mode-related effects are generally small (Tourangeau et al., 2013) and limited to socially desirable variables, such as those related to sensitive topics (e.g., health or income), for which PWPs may even yield more accurate responses due to reduced social desirability bias. The relatively low bias observed for Tourism was somewhat unexpected, although it may be attributed to the specific nature of the selected variables because some key variables (i.e., questions on domestic travel) within this survey demonstrated substantial discrepancies (see Bučar, 2022).

Overall, 205 of the 651 variables (31%) in the PWP had *RB'* values exceeding 10%, while the overall median *RB'* was 16%. This median is relatively high compared to the other evaluations presented in Section 2.2, where the median values ranged from 4.7% to 13%. The primary explanation for this discrepancy is probably the specific characteristics of the TPSs used in this study. Unlike most other surveys, which primarily covered health or general social indicators, the TPS sets we analyzed focused on core behavioral variables regarding official statistics.

The results also revealed that PWP respondents constituted a specific segment of the population characterized by lower levels of life satisfaction, social trust, health-related awareness, poorer health (but with fewer disabilities), more rental accommodation, less home ownership, and generally lower social status, although more travel was observed. However, this segment has higher internet usage, digital literacy, education involvement, and employment levels.

4.3 Limitations and future research

There are certain technical limitations worth highlighting with respect to the survey of NSIs. Namely, 19 of 31 NSIs responded (61%), which could have introduced a potential bias if nonresponding NSIs differed systematically regarding the PWP engagement. NSIs with fewer resources or lower PWP involvement may have chosen not to respond. In addition, although the sample included a comprehensive set of Eurostat-affiliated NSIs, generalizability to non-Eurostat NSIs remains limited. Future researchers should consider expanding the scope to encompass additional international statistical institutions. Moreover, incorporating more targeted items into the questionnaire may facilitate the identification of a wider range of technical, legal, operational, or cultural barriers to PWP implementation.

Several technical limitations also pertain to the analysis of *RB'* in estimates taken from PWPs. First, the PWPs were limited by 1KA panel characteristics, particularly the exclusion of internet nonusers. Integrating PWPs with mixed-mode approaches may enhance response rates (Callegaro et al., 2015; Scherpenzeel and Bethlehem, 2011; Vannieuwenhuyze, 2014) and reduce coverage bias (Eckman, 2016), especially among older or less technologically adept individuals. Second, recruitment relied solely on mailed invitations, which generally yield lower response rates than personal contact (Blom et al., 2017; Maslovskaya and Lugtig, 2022). Additionally, consent procedures were not optimized or experimentally varied (Bottoni and Sommer, 2019; Sakshaug et al., 2020). Future researchers should further investigate recruitment and consent strategies (Callegaro et al., 2015; Vehovar and Beullens, 2018). Third, the relatively small sample size in this PWP study (i.e., 1 628 panelists) may have reduced the quality of the estimates and compromised representativeness. Fourth, the omnibus structure of the PWP survey may have shaped the respondent context differently from that of corresponding single-topic TPS surveys, thereby constraining topic-specific generalizability. Fifth, the weighting model employed may have insufficiently accounted for unobserved confounders, such as attitudinal or behavioral traits, that influence PWP nonresponse or coverage bias (Mercer et al., 2018). In addition, the trimming of the weights above a threshold of 5 may have resulted in many extreme values. Future researchers should assess more advanced weighting strategies, including propensity score adjustments (Vehovar and Beullens, 2018) and adaptive survey designs that incorporate real-time error diagnostics (Schouten et al., 2017). Integrating topic-relevant auxiliary variables may further improve the accuracy of weighting procedures for future PWP applications (Callegaro et al., 2014). Sixth, transitioning from mailed vouchers to digital incentives may enhance cost efficiency. Finally, although attrition across the five PWP waves was minimal (one percentage point in each wave; see Section 4.1), both panel conditioning and attrition may have introduced data quality concerns (Kennedy et al., 2016).

In addition to technical issues, several conceptual aspects warrant discussion. Although we identified topic-specific variations in *RB'* (e.g., high *RB'* for ICT but low for Consumers), we did not assess the possible causal mechanisms of these differences. In this context, the high variation in the proportion of estimates with *RB'* > 10% (range: 0%–69% across 12 survey sets) indicates the need for a more granular understanding of the substantive and methodological factors affecting *RB'* levels. Separate treatment of nonresponse, noncoverage, and measurement errors would also provide additional insight (e.g., Bosnjak et al., 2018; Couper, 2011; Groves, 2006; Struminskaya et al., 2015; Vehovar and Beullens, 2018). Furthermore, we limited the *RB'* comparisons to cross-sectional estimates across PWP waves without assessing longitudinal trends or measurement equivalence (Cernat and Revilla, 2021; Cornesse and Blom, 2020). Future researchers should incorporate item-level diagnostics, response quality indicators, and error decomposition frameworks (e.g., Pennay et al., 2018; Unangst et al., 2020).

Furthermore, the bias-related findings are limited to a single PWP in Slovenia and may not be generalizable to contexts with different demographic, societal, or technological characteristics, although Slovenia is often considered a typical EU country, with median levels of socioeconomic indicators, including an internet penetration rate of 91% among individuals aged 16–74 years (Jevnikar, 2025). Similarly, the 651 variables we analyzed were purposively selected from 12 survey sets, although many more variables were considered in the corresponding TPSs. Although these selected variables were often essential for or typical of the corresponding TPSs, they may not be the most consistently representative. Ideally, a PWP would fully represent an entire TPS survey.

Nevertheless, despite these constraints, the 1KA Panel demonstrated favorable recruitment and retention outcomes, effectively functioning as a prototype for web probability panels in Slovenia. The cumulative response rate across all three recruitment channels and five PWP waves (2022–2023), based on a 9 100-unit sample, was 16%, aligning with the results reported by Villar et al. (2018) and Kocar and Kaczmirek (2024). This study thus represents a typical scenario in which PWPs (rather than TPSs) are used for official statistics. Increased resource allocation to improve 1KA Panel operations and procedures could potentially raise ORRs (e.g., above 20%) and reduce certain biases; however, it is likely that the overall pattern of the biases will remain largely unchanged (i.e., a considerable number of topics in official statistics seem to involve unacceptable biases when included in PWPs).

CONCLUSION

This article addresses a notable gap in the literature regarding the use, experience, and prospects of PWPs for official statistics. In this context, a survey conducted among NSIs in the EU provided a specific insight into the adoption of PWPs for official statistics. The related findings indicated that the adoption process is well underway, with one-quarter of NSIs already engaged in the PWP implementation. The results further revealed that, in addition to various methodological challenges that may increase the bias of estimates (e.g., attrition, conditioning, nonresponse, noncoverage, and measurement), a primary concern among NSIs was the lack of expertise and established good practices related to PWPs.

The other focus of this article was the bias of the PWP estimates. We employed a case study of Slovenian official statistics using the 1KA Panel to compare the PWP estimates with those from concurrently conducted TPS sets based on 651 variables taken from 12 official statistics surveys. The results revealed substantial discrepancies, with the PWP estimates exhibiting considerably more bias than TPS estimates. The median *RB'* across corresponding estimates was 16%, while the median proportion of variables with *RB'* exceeding 10% across the 12 survey sets was 31%. The PWP estimates demonstrated higher *RB'* for behavioral and factual variables, particularly for ICT usage (69% of estimates with *RB'* > 10%) and living conditions (SILC, 61%; Activity, 53%). In contrast, attitudinal variables exhibited lower *RB'*, such as for

Consumers (0%). In general, the biases identified for 651 variables suggest that PWP respondents have specific characteristics: more active lifestyles and internet usage but also lower social statuses.

Identifying bias patterns across survey topics allowed us to propose a framework for strategically allocating the use of TPSs to surveys with low or moderate *RB*' values. The findings suggest that TPSs may be unnecessary for surveys with consistently low *RB*' values, such as those focusing on attitudinal measures. In particular, our results suggest that PWPs may be suitable for producing certain estimates in surveys focused on general social attitudes, where relatively low bias levels were observed (e.g., in the Consumers and COVID survey sets). However, the potential cost savings of larger PWP samples may be insufficient to offset the higher *RB*' values observed in behavioral and factual domains. Therefore, despite the cost- and time-efficiency gains offered by PWPs, the extent of bias they involve remains a substantial limitation.

Thus, this article offers a general overview of PWP adoption for official statistics and provides a case study illustrating which survey topics are appropriate for PWPs as potential replacements for TPSs. Future researchers should address the limitations of empirical studies and refine their methodological approaches.

ACKNOWLEDGMENT

The Slovenian Research and Innovation Agency supported this research (grant numbers P5-0399, J5-3100, V5-2157, and J5-50159).

References

1KA PANEL. (2025). *1KA Panel Methodology* [online]. Ljubljana: Centre for Social Informatics. [cit. 22.9.2025]. <https://panel.1ka.si/?a=metodologija&lang_id=2>.

AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH. (2023). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* [online]. Alexandria, Virginia: AAPOR. [cit. 22.9.2025]. <https://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-edition.pdf>.

BACH, R. L., CORNESSE, C., DAIKELER, J. (2024). Equipping the Offline Population with Internet Access in an Online Panel: Does It Make a Difference? [online]. *Journal of Survey Statistics and Methodology*, 12(1): 80–93. <http://doi.org/10.1093/jssam/smad003>.

BERRENS, R. P., BOHARA, A. K., JENKINS-SMITH, H., SILVA, C., WEIMER, D. L. (2003). The Advent of Internet Surveys for Political Research: a Comparison of Telephone and Internet Samples [online]. *Political Analysis*, 11(1): 1–22. <http://doi.org/10.1093/pan/11.1.1>.

BETHLEHEM, J. (2014). Web Surveys in Official Statistics. In: ENGEL, U., JANN, B., LYNN, P., SCHERPENZEEL, A., STURGIS, P. (eds.). *Improving Survey Methods*, New York: Routledge.

BLOM, A. G., HERZING, J. M. E., CORNESSE, C., SAKSHAUG, J. W., KRIEGER, U., BOSSERT, D. (2017). Does the Recruitment of Offline Households Increase the Sample Representativeness of Probability-Based Online Panels? Evidence From the German Internet Panel [online]. *Social Science Computer Review*, 35(4): 498–520. <http://doi.org/10.1177/0894439316651584>.

BOSNJAK, M., DANNWOLF, T., ENDERLE, T., SCHAURER, I., STRUMINSKAYA, B., TANNER, A., WEYANDT, K. W. (2018). Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany: The GESIS Panel [online]. *Social Science Computer Review*, 36(1): 103–115. <http://doi.org/10.1177/0894439317697949>.

BOTTONI, G. (2023). *CROss-National Online Survey (CRONOS-2) Panel: Data and Documentation User Guide* [online]. London: ESS ERIC. [cit. 22.9.2025]. <https://www.europeansocialsurvey.org/sites/default/files/2023-07/CRONOS2_user_guide_e01_0.pdf>.

BOTTONI, G., SOMMER, E. (2019). *Best Strategies to Recruit and Maintain Web Panels* [online]. London: Synergies for Europe's Research Infrastructures in the Social Sciences. [cit. 22.9.2025]. <https://seriss.eu/resources/deliverables/>.

BUČAR, N. (2022). *Stroški in Napake v Anketi Turistična Potovanja Domačega Prebivalstva* [online]. Ljubljana: University of Ljubljana. [cit. 22.9.2025]. <https://repozitorij.uni-lj.si/IzpisGradiva.php?id=143434>.

CALLEGARO, M., BAKER, R. P., BETHLEHEM, J., GÖRITZ, A. S., KROSNICK, J. A., LAVRAKAS, P. J. (2014). (eds.). *Online Panel Research: A Data Quality Perspective*. Chichester: Wiley.

CALLEGARO, M., DISOGRA, C. (2008). Computing Response Metrics for Online Panels [online]. *Public Opinion Quarterly*, 72(5): 1008–1032. <https://doi.org/10.1093/poq/nfn065>.

CALLEGARO, M., LOZAR MANFREDA, K., VEHOVAR, V. (2015). *Web Survey Methodology*. London: SAGE.

CASEY, B. (2024). *Can We Trust the UK Labour Force Survey?* [online]. London: UK in a Changing Europe. [cit. 22.9.2025]. <https://ukandeu.ac.uk/can-we-trust-the-uk-labour-force-survey/>.
CENTERDATA. (2024). *LISS Panel* [online]. Tilburg: LISS Panel. [cit. 22.9.2025]. <https://www.lissdata.nl/>.
CERNAT, A., REVILLA, M. (2021). Moving From Face-to-Face to a Web Panel: Impacts on Measurement Quality [online]. *Journal of Survey Statistics and Methodology*, 9(4): 745–763. <http://doi.org/10.1093/jssam/smaa007>.
COCHRAN, W. G. (1977). *Sampling Techniques.* 3rd Ed. New York: Wiley.
CORNESSE, C., BLOM, A. G. (2020). Response Quality in Nonprobability and Probability-Based Online Panels [online]. *Sociological Methods & Research*. <http://doi.org/10.1177/0049124120914940>.
CORNESSE, C., BLOM, A. G., DUTWIN, D., KROSNICK, J. A., DE LEEUW, E. D., LEGLEYE, S., PASEK, J., PENNAY, D., PHILLIPS, B., SAKSHAUG, J. W., STRUMINSKAYA, B., WENZ, A. (2020). A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research [online]. *Journal of Survey Statistics and Methodology*, 8(1): 4–36. <http://doi.org/10.1093/jssam/smz041>.
CORNESSE, C., BOSNJAK, M. (2018). Is There an Association Between Survey Characteristics and Representativeness? A Meta-Analysis [online]. *Survey Research Methods*, 12(1): 1–13. <http://doi.org/10.18148/srm/2018.v12i1.7205>.
COUPER, M. P. (2011). The Future of Modes of Data Collection [online]. *Public Opinion Quarterly*, 75(5): 889–908. <http://doi.org/10.1093/poq/nfr046>.
COUPER, M. P. (2017). New Developments in Survey Data Collection [online]. *Annual Review of Sociology*, 43(1): 121–145. <http://doi.org/10.1146/annurev-soc-060116-053613>.
ECKMAN, S. (2016). Does the Inclusion of Non-Internet Households in a Web Panel Reduce Coverage Bias? [online]. *Social Science Computer Review*, 34(1): 41–58. <http://doi.org/10.1177/0894439315572985>.
EUROPEAN SOCIAL SURVEY. (2023). *CROss-National Online Survey (CRONOS-2) Panel* [online]. London: ESS ERIC. [cit. 22.9.2025]. <https://www.europeansocialsurvey.org/methodology/methodological-research/modes-data-collection/cronos>.
EUROPEAN SOCIAL SURVEY. (2025). *New Panel Survey Data Now Available* [online]. London: ESS ERIC. [cit. 22.9.2025]. <https://www.europeansocialsurvey.org/news/article/new-panel-survey-data-now-available>.
FITZGERALD, R. (2022). *ESS Electronic Questionnaire Device* [online]. London: ESS ERIC. [cit. 22.9.2025]. <https://www.ncrm.ac.uk/documents/EQD-presentation_rory%20fitzgerald.pdf>.
GOLINI, N., RIGHI, P. (2024). Integrating Probability and Big Non-Probability Samples Data to Produce Official Statistics [online]. *Statistical Methods & Applications*, 33(2): 555–580. <http://doi.org/10.1007/s10260-023-00740-y>.
GRAF, N. L., JONES, G., MILLER, T. (2025). Testing Self-Administration of the Consumer Expenditure Diary Survey in a Probability-Based Online Panel [online]. Los Angeles: *CIPHER 2025 Conference*. [cit. 22.9.2025]. <https://dornsife.usc.edu/cesr/wp-content/uploads/sites/54/2025/03/GRAF-main.pdf>.
GROVES, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys [online]. *Public Opinion Quarterly*, 70(5): 646–675. <http://doi.org/10.1093/poq/nfl033>.
GROVES, R. M., FOWLER, F. J., COUPER, M. P., LEPKOWSKI, J. M., SINGER, E., TOURANGEAU, R. (2009). *Survey Methodology*. 2nd Ed. New York: Wiley.
GROVES, R. M., PEYTCHEVA, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: a Meta-Analysis [online]. *Public Opinion Quarterly*, 72(2): 167–189. <http://doi.org/10.1093/poq/nfn011>.
HAYS, R. D., LIU, H., KAPTEYN, A. (2015). Use of Internet Panels to Conduct Surveys [online]. *Behavior Research Methods*, 47(3): 685–690. <http://doi.org/10.3758/s13428-015-0617-9>.
IVANOVSKA, A., BOSNJAK, M., VEHOVAR, V. (2025). Relative Bias in Health Estimates From Probability-Based Online Panels: Systematic Review and Meta-Analysis [online]. *Journal of Health and Rehabilitation Sciences*, 4(1): 10–23. <http://doi.org/10.33700/jhrs.4.1.137>.
JABKOWSKI, P., KOŁCZYŃSKA, M. (2020). Sampling and Fieldwork Practices in Europe: Analysis of Methodological Documentation From 1,537 Surveys in Five Cross-National Projects, 1981–2017 [online]. *Methodology*, 16(3): 186–207. <http://doi.org/10.5964/meth.2795>.
JEVNIKAR, Ž. (2025). *Usage of Internet in Households and by Individuals* [online]. Ljubljana: Statistical Office of the Republic of Slovenia. [cit. 22.9.2025]. <https://www.stat.si/StatWeb/en/News/Index/13336>.
KENNEDY, C., MERCER, A. W., KEETER, S., HATLEY, N., MCGEENEY, K., GIMENEZ, A. (2016). *Evaluating Online Nonprobability Surveys* [online]. Washington, DC: Pew Research Center. [cit. 22.9.2025]. <https://www.pewresearch.org/wp-content/uploads/2016/04/Nonprobability-report-May-2016-FINAL.pdf>.
KOCAR, S., BIDDLE, N. (2023). Do We Have to Mix Modes in Probability-Based Online Panel Research to Obtain More Accurate Results? [online]. *Methods, Data, Analyses*, 17(1): 93–120. <http://doi.org/10.12758/mda.2022.11>.
KOCAR, S., KACZMIREK, L. (2024). A Meta-Analysis of Worldwide Recruitment Rates in 23 Probability-Based Online Panels, Between 2007 and 2019 [online]. *International Journal of Social Research Methodology*, 27(5): 589–604. <http://doi.org/10.1080/13645579.2023.2242202>.
KOCH, A., FITZGERALD, R., STOOP, I., WIDDOP, S. (2010). *Field Procedures in the European Social Survey Round 5: Enhancing Response Rates* [online]. London: ESS ERIC. [cit. 22.9.2025]. <https://www.academia.edu/88218012/Field_Procedures_in_the_European_Social_Survey_Round_5_Enhancing_Response_Rates>.

LEMCKE, J., LOSS, J., ALLEN, J., ÖZTÜRK, I., HINTZE, M., DAMEROW, S., KUTTIG, T., WETZSTEIN, M., HÖVENER, C., HAPKE, U., ZIESE, T., SCHEIDT-NAVE, C., SCHMICH, P. (2024). Health in Germany: Establishment of a Population-Based Health Panel [online]. *Journal of Health Monitoring*, 9(S2): 1–21. <http://doi.org/10.25646/11992.2>.

MASLOVSKAYA, O., LUGTIG, P. (2022). Representativeness in Six Waves of CROss-National Online Survey (CRONOS) Panel [online]. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. <http://doi.org/10.1111/rssa.12801>.

MASLOVSKAYA, O., LYNN, P., CALDERWOOD, L., DURRANT, G., FITZGERALD, R., NICOLAAS, G., WILLIAMS, J. (2025). *Response Rates Position Statement* [online]. Colchester: Survey Futures. [cit. 22.9.2025]. <https://surveyfutures.net/publications/response-rates-position-statement/>.

MERCER, A., LAU, A. (2023). *Comparing Two Types of Online Survey Samples* [online]. Washington, DC: Pew Research Center. [cit. 22.9.2025]. <https://www.pewresearch.org/methods/2023/09/07/comparing-two-types-of-online-survey-samples>.

MERCER, A. W., LAU, A., KENNEDY, C. (2018). *For Weighting Online Opt-In Samples, What Matters Most?* [online]. Washington, DC: Pew Research Center. [cit. 22.9.2025]. <https://www.pewresearch.org/methods/2018/01/26/for-weighting-online-opt-in-samples-what-matters-most/>.

OECD. (2014). *PIAAC Technical Standards and Guidelines* [online]. Paris: OECD. [cit. 22.9.2025]. <https://www.oecd.org/en/about/programmes/piaac/piaac-data.html>.

OFFICE FOR NATIONAL STATISTICS. (2025). *Labour Force Survey Performance and Quality Monitoring Reports* [online]. Newport: Office for National Statistics. [cit. 22.9.2025]. <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports>.

OFFICE OF MANAGEMENT AND BUDGET. (2016). *Questions and Answers When Designing Surveys for Information Collections* [online]. Washington, DC: Office of Management and Budget. [cit. 22.9.2025]. <https://obamawhitehouse.archives.gov/sites/default/omb/inforeg/pmc_survey_guidance_2006.pdf>.

OLSON, K., WAGNER, J., ANDERSON, R. (2021). Survey Costs: where Are We and what Is the Way forward? [online]. *Journal of Survey Statistics and Methodology*, 9(5): 921–942. <http://doi.org/10.1093/jssam/smaa014>.

PASEK, J. (2016). When Will Nonprobability Surveys Mirror Probability Surveys? Considering Types of Inference and Weighting Strategies as Criteria for Correspondence [online]. *International Journal of Public Opinion Research*, 28(2): 269–291. <http://doi.org/10.1093/ijpor/edv016>.

PASEK, J. (2022). *Anesrake: ANES Raking Implementation* [online]. R Foundation for Statistical Computing. <http://doi.org/10.32614/CRAN.package.anesrake>.

PENNAY, D., 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* [online]. Canberra: ANU Centre for Social Research and Methods. [cit. 22.9.2025]. <https://csrm.cass.anu.edu.au/research/publications/online-panels-benchmarking-study-total-survey-error-comparison-findings>.

PEYTCHEV, A. (2025). *Evaluating Error From the Addition of Web Panel Data to a National Probability-Based Interviewer-Administered Survey* [online]. Washington, D.C.: CIPHER 2025 Conference. [cit. 22.9.2025]. <https://dornsife.usc.edu/cesr/wp-content/uploads/sites/54/2025/03/3-Peytchev.pdf>.

PFORR, K., DANNWOLF, T. (2017). What Do We Lose with Online-Only Surveys? Estimating the Bias in Selected Political Variables Due to Online Mode Restriction [online]. *Statistics, Politics and Policy*, 8(1): 105–120. <http://doi.org/10.1515/spp-2016-0004>.

R CORE TEAM. (2024). *The R Project for Statistical Computing* [online]. R Foundation for Statistical Computing. [cit. 22.9.2025]. <https://www.R-project.org>.

SAKSHAUG, J. W., HÜLLE, S., SCHMUCKER, A., LIEBIG, S. (2020). Panel Survey Recruitment with or without Interviewers? Implications for Nonresponse, Panel Consent, and Total Recruitment Bias [online]. *Journal of Survey Statistics and Methodology*, 8(3): 540–565. <http://doi.org/10.1093/jssam/smz012>.

SAKSHAUG, J. W., KROSNICK, J. A., ACKERMANN, A., MALKA, A., DEBELL, M., TURAKHIA, C. (2009). Dispositions and Outcome Rates in the 'Face-to-Face Recruited Internet Survey Platform' (the FFRISP). *AAPOR 64th Annual Conference,* Hollywood, Florida.

SÄRNDAL, C.-E., SWENSSON, B., WRETMAN, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

SCHERPENZEEL, A., BETHLEHEM, J. (2011). How Representative Are Online Panels? Problems of Coverage and Selection and Possible Solutions. In: DAS, M., ESTER, P., KACZMIREK, L. (eds.). *Social and Behavioral Research and the Internet*, New York: Routledge, pp. 105–132.

SCHOUTEN, B., PEYTCHEV, A., WAGNER, J. (2017). *Adaptive Survey Design*. New York: CRC Press.

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 [online]. *Asian Journal for Public Opinion Research*, 11(3): 257–273. <http://doi.org/10.15206/ajpor.2023.11.3.257>.

SMITH, P. (2022). *What Is Knock-to-Nudge and Does It Have a Future?* [online]. London: Ipsos. [cit. 22.9.2025]. <https://www.ncrm.ac.uk/documents/4_Patten%20Smith_Data%20collection%20network%202_2_22.pdf>.

ŠOŠTARIČ, T. (2020). *Izzivi Spletnih Verjetnostnih Panelov: Primer Panela CRONOS* [online]. Ljubljana: University of Ljubljana. [cit. 22.9.2025]. <https://repozitorij.uni-lj.si/Dokument.php?id=135272&lang=slv>.

ŠTRLEKAR, L., VEHOVAR, V. (2025). *SurveyWeightingGUI* [online]. Ljubljana: 1KA Panel. [cit. 22.9.2025]. <https://enklikanketa.shinyapps.io/SurveyWeightingGUI/>.

STRUMINSKAYA, B., BOSNJAK, M. (2021). Panel Conditioning: Types, Causes, and Empirical Evidence of what We Know So Far [online]. In: LYNN, P. (eds). *Advances in Longitudinal Survey Methodology*, New York: Wiley, pp. 272–301. <http://doi.org/10.1002/9781119376965.ch12>.

STRUMINSKAYA, B., DE LEEUW, E., KACZMIREK, L. (2015). Mode System Effects in an Online Panel Study: Comparing a Probability-Based Online Panel with Two Face-to-Face Reference Surveys [online]. *Methods, Data, Analyses*, 9(1): 3–56. <http://doi.org/10.12758/mda.2015.001>.

SVENSSON, J. (2014). Web Panel Surveys – a Challenge for Official Statistics [online]. *Proceedings of Statistics Canada Symposium*, Ottawa: Statistics Canada. [cit. 22.9.2025]. <https://www.statcan.gc.ca/eng/conferences/symposium2014/program/14280-eng.pdf>.

TOURANGEAU, R., BRICK, J. M., LOHR, S., LI, J. (2017). Adaptive and Responsive Survey Designs: a Review and Assessment [online]. *Journal of the Royal Statistical Society*, 180(1): 203–223. <http://doi.org/10.1111/rssa.12186>.

TOURANGEAU, R., CONRAD, F., COUPER, M. (2013). *The Science of Web Surveys*. New York: Oxford University Press.

UNANGST, J., AMAYA, A. E., SANDERS, H. L., HOWARD, J., FERRELL, A., KARON, S., 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 [online]. *Journal of Survey Statistics and Methodology*, 8(1): 62–88. <http://doi.org/10.1093/jssam/smz040>.

U.S. CENSUS BUREAU. (2025). Agency Information Collection Activities; Submission to the Office of Management and Budget (OMB) for Review and Approval; Comment Request; High-Frequency Surveys Program/Household Trends and Outlook Pulse Survey (HTOPS) [online]. Washington, DC, *Federal Register*, 90(39): 10879. [cit. 22.9.2025]. <https://www.federalregister.gov/documents/2025/02/28/2025-03194/agency-information-collection-activities-submission-to-the-office-of-management-and-budget-omb-for>.

VANNIEUWENHUYZE, J. (2014). On the Relative Advantage of Mixed-Mode Versus Single-Mode Surveys [online]. *Survey Research Methods*, 8(1): 31–42. <http://doi.org/10.18148/srm/2014.v8i1.5500>.

VEHOVAR, V., BEULLENS, K. (2018). Cross-National Issues in Response Rates [online]. In: VANNETTE, D. L., KROSNICK, J. A. (eds.). *The Palgrave Handbook of Survey Research*, Cham: Springer International Publishing, pp. 29–42. <http://doi.org/10.1007/978-3-319-54395-6_6>.

VEHOVAR, V., ČEHOVIN, G., PRAČEK, A. (2023). *The Use of Probability Web Panels in National Statistical Institutes* [online]. Ljubljana: Centre for Social Informatics. [cit. 22.9.2025]. <https://1ka.cdi.si/uploadi/editor/doc/1676613266NSIReport-final.pdf>.

VEHOVAR, V., TOEPOEL, V., STEINMETZ, S. (2016). Non-Probability Sampling. In: WOLF, C., JOYE, D., SMITH, T. W., FU, Y. (eds.). *The SAGE Handbook of Survey Methodology*, London: SAGE, pp. 329–346.

VILLAR, A., SOMMER, E., FINNØY, D., GAIA, A., BERZELAK, N., BOTTONI, G. (2018). *Cross-National Online Survey (CRONOS) Panel: Data and Documentation User Guide* [online]. London: ESS ERIC. [cit. 22.9.2025]. <https://openaccess.city.ac.uk/id/eprint/30024/1/CRONOS_user_guide_e01_1.pdf>.

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 [online]. *Public Opinion Quarterly*, 75(4): 709–747. <http://doi.org/10.1093/poq/nfr020>.

APPENDIX

The dataset corresponding to the 651 estimates we analyzed is available at:
<https://doi.org/10.5281/zenodo.15829184>.