

Centre for Social Informatics
Faculty of Social Sciences
University of Ljubljana

Evaluating probability and nonprobability online panels

Vehovar Vasja, Gregor Čehovin

Ljubljana, March 2026

Working papers from the *Centre for Social Informatics* are intended for analysis, discussion and commentary. They are not peer reviewed and do not reflect the views of the *Centre*, the *Faculty* or the *University*. The authors are solely responsible for any errors in the working papers.

© 2026 *Centre for Social Informatics*. All rights reserved. Excerpts from working papers not exceeding two paragraphs may be cited in other publications without special permission, provided that full information about the authors is included. For working papers commissioned by an external client, the client reserves full rights to decide about future publications.

Below is a summary of findings on evaluation of probability-based web panels (PWP) and nonprobability web panels (NWP). Empirical research was conducted in Slovenia with 1,034 items (i.e., variables) from 11 survey sets (Appendix, Table A1) implemented in three research settings:

- PWP:** largest national probability panel *IKA* (overall response rate 17%) <https://panel.ika.si/en/>
- NWP:** largest national access panel, *Valicon Opinia Club*, <https://opinia.club/gb-eng/>,
- TPSs** (traditional probability surveys): *Statistical Office* and *National Health Institute* surveys based on F2F or P2W with traditional follow-ups (overall response rates above 50%). The *European Social Survey*, R8, Slovenian implementation, was also added (F2F, response rate 54%). PWP and NWP have sample sizes around $n=1500$, while TPSs were large official and health surveys.

Finding 1: Expectedly, the estimates (of course, after proper weighting) from PWP differed considerably from TPSs. Median relative bias (*RB*; definition in Appendix) was 14% (Table 1). The share of items with statistically significant bias and estimated *RB*s above 10% was also high (30%) but strongly varied, from tourism (8%) to ICT (69%). The ESS module, the only survey set focused on attitudes (behavioral and factual variables prevail in official/health surveys), also showed same median *RB* (14%) and high (27%) share of variables with statistically significant bias and *RB*s above 10%.

Finding 2: Contrary to both theoretical predictions and our anticipated outcomes, the differences between PWP and NWP estimates were unexpectedly small. The median *RB* for PWP-TPS was 14%, while it was 16% for NWP-TPS (Table 1). The overall share of variables with *RB*s above 10% and statistically significant bias was 30% for PWP and 35% for NWP. The direct NWP-PWP comparison showed a median *RB* of 11% (meaning that 50% of all estimates in NWP had *RB* above 11%), while 15% of estimates had *RB*s above 10% and statistically significant bias. For the ESS module, the NWP-PWP differences were also relatively small (10% and 11% respectively).

Table 1. Paired comparisons across PWP, NWP, and TPS implementations: item (variable) counts, median *RB*, share of items with statistically significant bias ($p < .05$) and *RB* exceeding 10% or 20%

Survey set	Item count	PWP vs. TPS			NWP vs. PWP			NWP vs. TPS		
		Median <i>RB</i> (%)	<i>RB</i> > 10%	<i>RB</i> > 20%	Median <i>RB</i> (%)	<i>RB</i> > 10%	<i>RB</i> > 20%	Median <i>RB</i> (%)	<i>RB</i> > 10%	<i>RB</i> > 20%
Activity	17	29	53%	41%	8	0%	0%	46	53%	47%
COVID	58	11	14%	10%	7	4%	2%	10	8%	6%
Drugs	141	20	21%	15%	13	10%	6%	19	27%	22%
ESS “Core”	392	14	27%	17%	10	11%	8%	21	35%	24%
ICT (2023)	39	22	69%	51%	10	13%	5%	16	59%	36%
ICT (2024)	168	9	25%	14%	6	17%	11%	10	37%	26%
Mental health	68	10	21%	13%	3	6%	4%	9	7%	3%
SILC	57	23	61%	49%	18	47%	37%	56	86%	79%
SILC (HH)	54	14	57%	35%	14	43%	33%	14	50%	37%
Tourism (persons)	24	7	8%	4%	12	42%	13%	10	33%	29%
Tourism (HH)	16	29	44%	38%	15	19%	13%	26	50%	44%
All survey sets	1,034	14	30%	20%	11	15%	10%	16	35%	25%

Comparisons with Pew Research: A conceptually identical study in the US (see Appendix, Table A2) showed median absolute *RB*s of 12% (PWP-TPS) and 29% (NWP-TPS). The US PWP thus performed similarly as reported above (despite much lower overall response rates for PWP in the US: 1.4%, 3% and 7%), while NWP in the US performed substantially worse. Of course, the comparisons with Pew Research serve as an observational illustration due to specific survey conditions in the US.

Comparisons with Slovenian CRONOS PWP: The European Social Survey (ESS) recruited its PWP (i.e., CRONOS) via follow-up to the main ESS survey (F2F). We compared the main ESS with the subset who later participated in CRONOS (around 40% of main ESS respondents). The main ESS served as the benchmark (TPS), while (additionally reweighted) estimates from the subset served as a proxy for PWP estimates. The PWP proxy estimates were thus also based on the main ESS survey (i.e., F2F). Therefore, only attrition (i.e. nonresponse) affected the comparisons (no mode effects, noncoverage, or other factors present in real-life PWPs). The three CRONOS rounds (ESS-8, 10, 11) for Slovenia showed a median *RB* of 4.5% (PWP-TPS) and 7% of estimates with statistically significant bias and *RB* exceeding 10%. These values are much smaller compared to Table 1 (PWP-TPS: 14% and 27%). This is mainly due to the above-described dependent samples, but in part also to the higher overall response rate in CRONOS (24%), additional postal mail option for internet nonusers (*IKA panel*, PWP in our study used only web, excluding internet nonusers - internet penetration was 88%) and much higher overall budget in ESS CRONOS (including higher incentives).

Comparisons with all 26 CRONOS implementations (all countries across the three rounds) showed that the average share of variables with statistically significant bias and *RB*s above 10% was 9% (maximum 14%). These all are, of course, ESS-specific illustrations, conceptually different from the figures in Table 1. Nonetheless, they provide a valuable complementary perspective.

Finding 3: Linear scaling of respondents' characteristics for TPS-PWP-NWP: Compared to TPSs, the respondents in PWP are less happy, less trusting, and less religious. They travel more and are more technologically proficient, but show higher material, social, and health deprivation. They are less likely to be permanently employed and more often live in multi-apartment buildings. They are politically more active and left leaning, but are less favorable towards EU, migration (also less likely to be of immigrant origin) and responsibility for climate change. These characteristics were even more pronounced among NWP respondents (e.g., even less happy). When some of these survey sets were implemented with river sampling (e.g., advertising on Facebook and Instagram), the respondents were further profiled in the same direction. Again, these characteristics are specific to Slovenia, which may differ from typical characteristics in the literature.

Finding 4: The cost savings from PWP could not compensate for their lower accuracy compared to TPS. Namely, the accuracy (MSE) in PWP-TPS comparisons was dominated by bias (due to large sample sizes). The exceptions were surveys (or subgroups) with small sample sizes and items with relatively large standard deviations (e.g. small shares). Only in those situations could the increased sample size (based on cost savings) compensate for the lower accuracy of PWP relative to TPS.

Finding 5: Criteria for choosing between PWP and NWP. The differences between PWP and NWP were relatively small (although non-negligible). This raises question: **Why would one pay more and wait longer for a PWP when data of comparable quality can be obtained faster and at much lower cost in NWP?** Opinion, market and commercial surveys have already moved to NWPs, while some surveys (i.e. official statistics) strictly require TPS (at least P2W designs with expensive traditional follow-ups or knock-to-nudge efforts). Nevertheless, some researchers and clients are still choosing between PWP and NWP. To provide cognitive insight into this decision process, 22 in-depth interviews were conducted with researchers of various profiles from Slovenia and abroad. The following key attributes were identified: the anticipated level of relative bias (and its rate of recurrence), costs (and related sample size), budget level, budget flexibility, and the nature of the project (important policy-creation vs. general exploratory analysis). Results suggest that the decision-making process often relies on cognitive heuristics, with researchers setting a threshold for acceptable *RB* (e.g. 5%, or 10%). To further quantify decision-making thresholds, a conjoint-based questionnaire has been designed to present respondents with hypothetical scenarios requiring a selection between two options.

APPENDIX

Relative bias: Instead of absolute percentage differences between PWP or NWP survey estimates and benchmarks, the **absolute value of relative bias** (*RB*) is used here, expressing the difference as a proportion of the benchmark. For example, the same 2-percentage-point absolute bias has a very different interpretation with benchmarks 5% or 50% (e.g. estimates 7% and 52% respectively). *RB*s properly reflect these effects, as we have $RB_1 = 2\%/5\% = 0.40$ vs $RB_2 = 2\%/50\% = 0.04$. The benchmark values are typically from TPSs, but sometimes (as in the Pew Research study) also from administrative data. When PWP and NWP are compared, the PWP estimates are assumed as benchmark values.

Table A1. Overview of the 11 survey sets from TPS and replicated in the PWP and NWP

#	Survey set acronym: Full survey set name	Panel wave	Var. count	TPS provider	TPS mode
1	Activity: Active and inactive population (LFS)	4	17	SORS	CAPI, CATI
2	COVID: Work and living conditions	0	58	SORS	web, CATI
3	Drugs: National survey on tobacco, alcohol, drugs	2	141	NIPH	web, CAPI
4	ESS: Slovene Public Opinion (ESS “Core”)	3	392	FSS	CAPI
5	ICT 2023 <u>use</u> in households and by individuals	2	39	SORS	web, CAPI
6	ICT (2024) use in households and by individuals	5	168	SORS	web, CAPI
7	Mental health: Attitudes toward mental health	4	68	NIPH	web, PAPI
8	SILC: Living conditions survey (persons)	1	57	SORS	CAPI, CATI
9	SILC (HH): Living conditions survey (households)	1	54	SORS	CAPI, CATI
10	Tourism (persons): Tourist trips – domestic population	2	24	SORS	web, CATI
11	Tourism (HH): Tourist trips – domestic population	2	16	SORS	web, CATI
All survey sets			1,034		

Note: SORS = Statistical Office, NIPH = National Institute of Public Health of Slovenia, FSS=Faculty of Social Sciences.

Table A2. Calculations based on Pew Research data: median *RB*s for PWP and NWP

Survey set	PWP vs. TPS	NWP vs. TPS
	Median <i>RB</i> (%)	Median <i>RB</i> (%)
Consumer behavior	11	48
Family	5	16
Health	13	20
Living conditions and background variables	13	21
National politics	40	83
Work	7	24
All survey sets (231 variables)	12	29

Note: All Pew Research estimates were based on averages calculated from three different PWPs and three different NWP. For original data, see Mercer, A., & Lau, A. (2023); <https://www.jstor.org/stable/resrep63520>.

The Research Projects: The analyses presented were conducted as part of the project [J5-50159](#) (Assessing and Modelling Unit Nonresponse Bias in Probability-Based Web Surveys, October 1, 2023 – September 30, 2026; [Abstract](#)), which extends the earlier project [J5-3100](#) (Evaluating probability and nonprobability online panels October 1, 2021 – September 30, 2025; [Abstract](#)). Both projects are funded by the Slovenian Research and Innovation Agency ([ARIS](#)).