

ARTICLES

Breakoffs in an hour-long, online survey

Tom Emery¹, Susana Cabaco², Luisa Fadel³, Peter Lugtig⁴, Vera Toepoel⁵, Almut Schumann⁶, Detlev Lück⁶, Martin Bujard⁶

¹ Department of Public Administration & Sociology, Erasmus University Rotterdam, ² RIVM, the Netherlands, ³ UC Louvain, Belgium, ⁴ Utrecht University, the Netherlands, ⁵ Statistics Netherlands, the Netherlands, ⁶ Federal Institute for Population Research, Germany

Keywords: breakoffs, online survey, cross-national, Generations & Gender Programme

<https://doi.org/10.29115/SP-2023-0008>

Survey Practice

Vol. 16, Issue 1, 2023

We analyze paradata from an online pilot of the Generations and Gender Programme (GGP) conducted in three countries (Croatia, Germany, and Portugal) to understand the extent, timing, and patterns in breakoffs during a long online survey. The GGP is notable as an online survey given that the median length of a face-to-face interview is 52 minutes, and the survey was initially designed for face-to-face. Paradata was collected for 3,378 web surveys. Breakoffs before the questionnaire was completed occurred in 17% of these surveys. The analysis uses Cox regression models to explore the timing of breakoffs and the influence of contextual factors. The results indicate that the breakoff hazard does not increase or decrease across the length of the questionnaire. The risk of breakoff does vary considerably across countries, between genders, and also by partnership status. Respondents are twice as likely to breakoff on a loop question, and respondents completing the survey on a smartphone are 2.6 times as likely to breakoff as those using a tablet or PC. Respondents receiving a conditional incentive were 65% less likely to breakoff than those who did not. The lessons from this work can help inform future strategies converting existing long, cross-national face-to-face studies into an online format.

Introduction

A central tenant of designing online surveys is to keep it short. However, surveys that are transitioning to an online context generally have an exogenous number of items that need to be collected, limiting the degree to which a survey can be shortened. This challenge is particularly acute for large cross-national surveys such as the European Value Study (EVS), the European Social Survey (ESS), or the Generations and Gender Programme (GGP) (Wolf et al. 2021). Large cross-national surveys are much longer than general online surveys that have been the basis of previous analyses of breakoff rates and their impact on data quality (Couper and Peterson 2017; Galesic and Bosnjak 2009; Lambert and Miller 2015; Peytchev 2009; Steinbrecher, Rossmann, and Blumenstiel 2015; Toepoel and Lugtig 2014).

Existing literature on breakoff rates rarely examine a survey of such length (see [Table 1](#)), and it is unknown whether breakoff rates beyond 30 minutes remain steady, accelerate or decline, and whether this is dependent on the context of the surveys such as the device type used, incentives used or time of day of starting the survey. It is therefore difficult to assess the impact of breakoffs on data quality based on existing findings which focus on very short survey lengths (Chen, Cernat, and Shlomo 2023; Mittereder and West 2022).

We attempt to address this shortcoming by examining the level of breakoffs in a “Push-to-Web” (P2W) pilot of around 300 items and which takes approximately 50 minutes to complete, conducted by the GGP. We also consider the subsequent data quality considerations of transitioning online. Further information on the GGP, including technical guidelines, the questionnaires used in the pilot, and information on the questionnaire’s history and design process can be found on www.ggp-i.org. Specifically, we are using data from the GGP: Push-to-Web pilot which was conducted in 2018 (Lugtig et al. 2022).

The GGP has several characteristics that are commonly observed in large, comparative surveys that complicate the transition to predominantly online data collection and the impact of breakoffs on data quality (Gauthier and Emery 2014). First, the GGP is a long survey (Gauthier, Cabaço, and Emery 2018). In face-to-face settings, the interview averaged 52 minutes with a high degree of variability depending on the complexity of the respondents’ background. Second, the GGP is a cross-national survey and needs to ensure cross-national comparability despite differences in the uptake of online surveys and different adoption rates of various internet devices such as smartphones and tablets. Third, the GGP is sequenced to start with the most substantively important modules first. This means that the impact of a breakoff after 10 minutes differs markedly from one at 35 minutes. There is therefore not only concern with the completion rate but also the specific point in the questionnaire at which a breakoff occurred. Finally, the GGP is a long-standing survey and the items that are included are exogenous, and the survey was not initially composed with a web implementation in mind.

Existing research has regularly conflated breakoff rate and completion rate as antonyms. In this analysis, we argue that the specific timing of breakoff is particularly informative in understanding the impact on data quality and use a Cox proportional hazard model to assess the impact of survey design and survey context on when individuals breakoff. We contribute evidence on the impact of breakoffs on data quality in very long online surveys and outline strategies for mitigating the impact of breakoffs on transitioning a long survey online.

Data and Methods

The GGP Pilot Data

The GGP pilot study was conducted between May and November 2018 with a slight variation in fieldwork windows across countries due to fieldwork constraints by agencies. The data was collected using Blaise 5.3 on a single server hosted by Netherlands Interdisciplinary Demographic Institute (NIDI). Fieldwork management and contacts were managed by national teams in collaboration within country fieldwork agencies. Respondents were invited to take part in an initial invitation along with an unconditional incentive of €5 (40 kuna in Croatia) and then were given 6 weeks to complete the online

questionnaire. Respondents received a reminder 2 weeks after fieldwork started and then a further reminder 4 weeks after fieldwork started. There were two deviations from this design. An experimental subgroup in Croatia received two reminders, one week apart to ascertain whether it increased response rates. In Germany, there were several subgroups which used a variable incentive structure including mixes of both conditional and unconditional incentives.

The gross samples were derived from national registers of named individuals (Germany and Croatia), and random route (Portugal) where enumerators provided a list of sampled addresses. Then, a subsample of one reference group (face-to-face mode) and two or more treatment groups (push-to-web design) was randomly selected from the same gross sample in each country. In this way, we obtained a total gross sample size of 18,494 individuals for the online version of the survey (Germany N=8,496; Croatia N=2,900; Portugal N=7,098). The response rate in Portugal was much lower than in Croatia and Germany due the random route approach and a high noncontact rate.

From the gross sample, 3,378 respondents started the web version of the survey, which represents the baseline of our analytical sample (Germany N=1,894; Croatia N=1,153; Portugal=331). For this experiment, paradata was collected at a keystroke level.

To assess the point where participants dropped out during the survey, we used the last question in the survey that was answered by each respondent who began the survey (i.e., the question reached at breakoff).

Independent variables

The independent variables were from both the pilot survey responses and from the survey paradata. Information about participants' age, gender, education level, area of residence, whether respondents are or are not in a relationship, are derived from the survey responses, while information about the screen size, and the device used during the survey were collected from the paradata.

The digital questionnaire is designed to allow for completion on desktop computer, laptop, tablet, smartphones, and other devices with small screens. The screen size of the device respondents used to complete the survey has been created for three width categories: small (from 0 to 575.98 pixels), medium (from 576 to 991.98 pixels), and large (from 992 to 2000 pixels). Device was included as a time-varying covariate and based on the device size of the most recent log-in event.

In addition to this, we included an indicator for a subset of respondents in Germany that received an additional conditional incentive. This was part of a wider incentive study, but the indicator is included here to estimate whether conditional incentives delay or prevent breakoff. We included an indicator that specified whether a question was sensitive or part of a loop structure. Sensitive questions were related to sexual practices such as whether the respondent had

Table 1. Studies on break-off rates

Study	Fieldwork	Break-off rate (\%)	Aver. duration (min.)
Peytchev 2009	2003	14.9	21
Peytchev 2009	2004	8.6	18
Galesic and Bosnjak 2009	2009	9.4	10
Galesic and Bosnjak 2009	2009	12.5	20
Galesic and Bosnjak 2009	2009	11.3	30
Steinbrecher, Rossmann, and Blumenstiel 2015	2010	10.6	25
Steinbrecher, Rossmann, and Blumenstiel 2015	2011	16.3	25
Steinbrecher, Rossmann, and Blumenstiel 2015	2011	19.8	25
Steinbrecher, Rossmann, and Blumenstiel 2015	2011	12.7	25
Toepoel and Lugtig 2014	2012	1.4	5-10
Couper and Peterson 2017	2012	14.9	15
Couper and Peterson 2017	2013	13.6	15
Couper and Peterson 2017	2014	14.2	15
Lambert and Miller 2015	2014	9.3	28

Own elaboration based on the literature cited

Multiple entries from the same publication represent when respondents were allocated to different designs randomly

had sex in the last four weeks, whether it was unprotected sex, and whether they were sterile. Loop questions referred to recording of all previous cohabiting partners, all children, and social support network members.

The age range of respondents was between 18 and 49 years. Gender was dichotomously coded (male as reference). Education level was coded using three categories: low, medium, and high. The low level includes respondents with International Standard Classification of Education (ISCED) 0, 1 and 2, the medium level includes respondents with ISCED 3 and 4, while the high level includes respondents with ISCED 5 and 6.

Results

The mean duration of completed surveys was 65 minutes, which was longer than the face-to-face surveys but was inclusive of pauses in the completion that were less than 20 minutes between questions. The median completion time was 59 minutes. The final breakoff rate at the end of the survey was 17.23%, which means that 82.77% finished the survey.

In Croatia, 89.16% of respondents who started the survey, finished it. This is a breakoff rate that is lower than all but four of the 14 studies listed in [Table 1](#). In Germany and Portugal, the rates were much higher at 20.33% and 21.75%, respectively.

Breakoffs were relatively linear across nearly 300 questions. There were noticeable problem points in the survey where breakoff was high, but there was no clear association with time and little sign of an acceleration or deceleration

Table 2. Descriptive variables by country

		Germany		Croatia		Portugal		Total	
		N	%	N	%	N	%	N	%
Education	Low	264	13.9	556	48.2	35	10.6	855	25.3
	Medium	615	32.5	283	24.5	117	35.3	1,015	30
	High	1,015	53.6	314	27.2	179	54.1	1,508	44.6
Device	Small	467	24.7	338	29.3	64	19.3	869	25.7
	Medium	117	6.2	51	4.4	19	5.7	187	5.5
	Large	1,310	69.2	764	66.3	248	74.9	2,322	68.7
Total		1,894	100	1,153	100	331	100	3,378	100

in breakoffs. After the first 100 items, breakoffs stood at 3%, 6%, and 9% for Croatia, Germany, and Portugal, respectively. Between 100 and 200, a further 4%, 8%, and 7% broke off, respectively.

There were obvious cross-national differences in breakoff trajectories. Portugal and Germany had much steeper breakoff gradients than Croatia. There were also several points in the survey where there were obvious breakoffs. These corresponded with the start of loops which cover detailed questions on previous partners, all children, all other household members, and a social network module.

The large drop at item 116 in Portugal and Germany corresponds with the start of the social network module. In observed surveys, respondents repeatedly complained about the nature of these social network questions and their repetitive structure. It is unclear why the social network module did not have this effect in Croatia, but it should be noted that the number of refusals and don't knows for these items was higher in Croatia, suggesting that there was an inclination to skip these items rather than drop-out.

The results for the Cox regression of question breakoff are presented in [Table 3](#). These are expressed as coefficients, not hazard rates. Hazard rates reported in the text are derived from one minus the exponent of the coefficient. In model 1, the conditional incentives which were used for a random subsample in Germany were associated with a 55% lower chance of breakoff on each question. This also suggests that overall breakoff rate for Germany is held up by these conditional incentives. After including these conditional incentives, Croatians were 72% less likely to break off than Germans, or conversely, Germans were 2.6 times as likely to break off as those in Croatia.

The results of device size show a nonlinear relationship between screen size and breakoff. The least likely to breakoff were medium devices, which broadly captures tablets. Those using a smartphone were 3.6 times as likely to break off from the survey as those using a tablet.

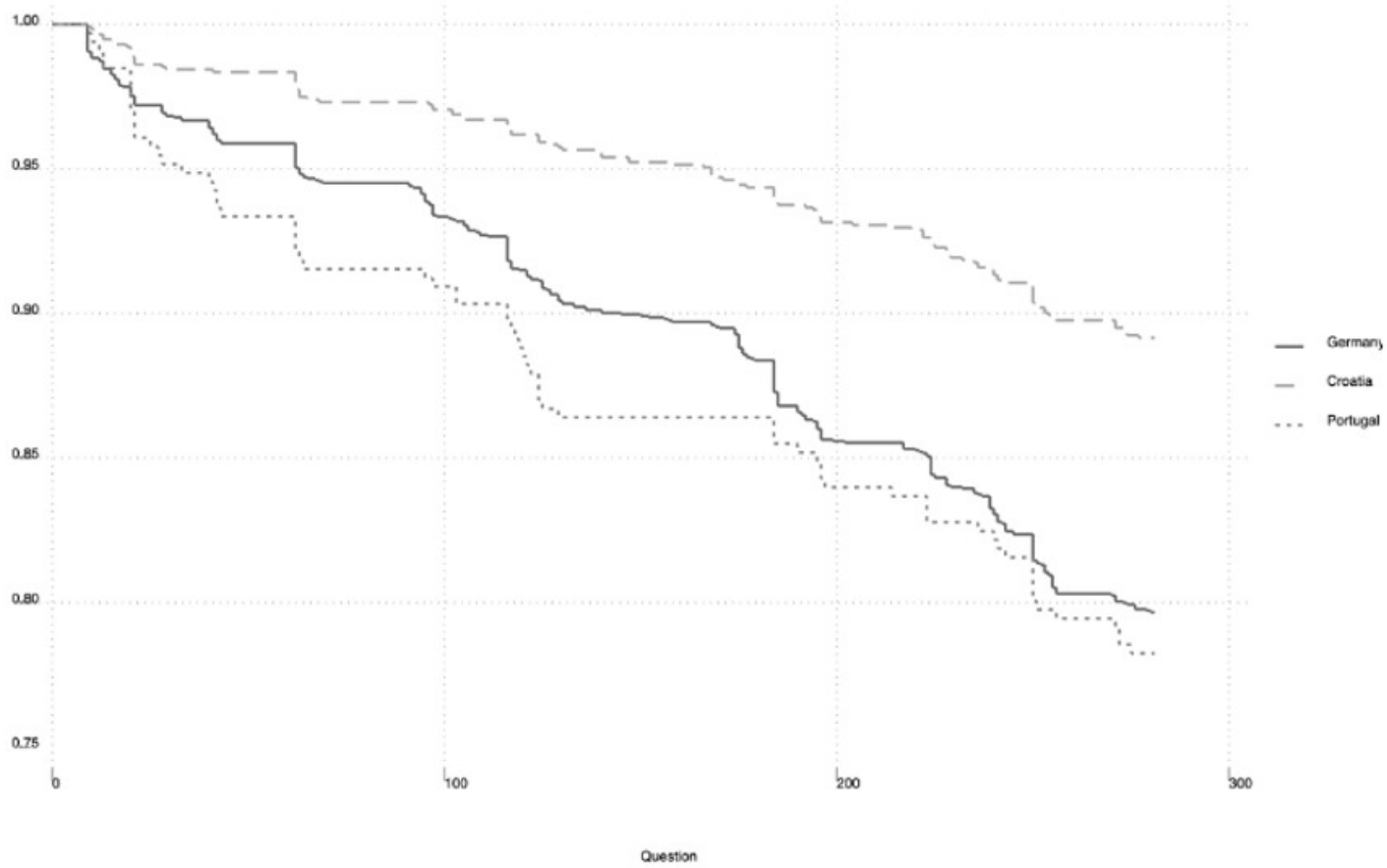


Figure 1

This is in the direction expected; however, the survival curves do exhibit specific points of deviation around items that are particularly problematic for smartphone devices such as the GGP social network module or extensive loops and suggest sharper breakoffs at the beginning. This is supported by the loop variable indicator which shows that across all respondents, they were twice as likely to break off when the variable was a loop variable as otherwise. An interaction effect between loop and device was included, but this was shown to be insignificant (results not shown). Sensitive questions about the respondents' sex life and fecundity were not associated with breakoff. This is perhaps because this was explicitly stated in the survey introduction.

Breakoff rates in Croatia were 71% lower than those in Germany. Croatia has a substantially lower internet penetration rate and lower average internet speed and yet the likelihood of breakoff in Croatia was much lower. This is particularly surprising given that some of the samples in Germany were provided with conditional and unconditional incentives while in Croatia all respondents were provided with unconditional incentives.

When we look at model 2, the results suggest that female respondents without a partner and male respondents with a partner are approximately as likely as each other to break off. This is, however, also true of the inverse, as female respondents with a partner appear to be as likely to break off as those men without a partner, with the latter pairing being the less likely to break off. We have no explanation for why single women and partnered men are more likely to break off.

Finally, we also tested to see whether any of the coefficients varied over time in models 5–8 in [Table 4](#). No interactions were statistically significant. We also replicated these results using a range of operationalizations of survey phase including separating into 2, 3, 5, and 10 sections, but this did not change the results. This supports the assertion from the survival curve that there is no observable trend in the breakoff rate over the length of the survey or over the impact of specific problematic variables such as loop variables.

Conclusions

An overall breakoff rate of 17% across a survey that takes almost an hour to complete and consists of nearly 300 questions, compares favorably with breakoff rates observed in shorter online surveys. It might be argued that low response rates in P2W compared to face-to-face compound the issue of breakoffs, but in the wider online experiment, P2W response rates were higher than face-to-face (Lutig et al. 2022). Survey researchers should reconsider the convention that shorter online surveys are always better, particularly in instances where the number of items is exogenous. If respondents are primed on the questionnaire length and motivated to participate, then breakoff itself may not be a sufficient reason not to pursue a longer questionnaire.

Table 3. Cox regression of Question Break Off

	1	2	3	4
Cond. Incentive	-0.806*** (-6.93)	-0.802*** (-6.90)	-0.805*** (-6.93)	-0.807*** (-6.93)
Screen: Medium	-1.316*** (-4.41)	-1.313*** (-4.40)	-1.042*** (-3.29)	-1.763*** (-3.06)
Screen: Large	-0.739*** (-7.38)	-0.739*** (-7.39)	-0.694*** (-5.76)	-0.784*** (-5.53)
Loop Question	0.754*** -2.21	0.755*** -2.21	0.754*** -2.21	0.754*** -2.21
Sensitive Question	-0.614 (-1.32)	-0.614 (-1.32)	-0.606 (-1.30)	-0.614 (-1.32)
Female	-0.113 (-1.18)	0.241*** -1.22	-0.122 (-1.26)	-0.113 (-1.18)
Partnered	0.145 -1.28	0.386*** -2.27	0.146 -1.28	0.146 -1.28
Edu: Medium	-0.442*** (-3.37)	-0.439*** (-3.36)	-0.445*** (-3.40)	-0.436*** (-3.33)
Edu: High	-0.507*** (-4.07)	-0.506*** (-4.07)	-0.506*** (-4.07)	-0.503*** (-4.05)
Rural	0.105 -1.08	0.113 -1.15	0.095 -0.97	0.04 -0.26
Country: HR	-1.259*** (-9.24)	-1.249*** (-9.18)	-1.243*** (-6.70)	-1.257*** (-9.22)
Country: PT	-0.208 (-1.13)	-0.203 (-1.10)	0.276 -0.91	-0.205 (-1.12)
Female X Partnered		-0.460*** (-2.03)		
Screen: Medium X Country: HR			-1.035 (-0.98)	
Screen: Medium X Country: PT			-4.62 (-6.23)	
Screen: Large X Country: HR			0.011 -0.05	
Screen: Large X Country: PT			-0.647 (-1.75)	
Screen: Medium X Rural				0.667 -0.99
Screen: Large X Rural				0.086 -0.44
Observations	3378	3378	3378	3378

DE = Germany, HR = Croatia, PT = Portugal

t-statistics in parentheses

* p < 0.05, ** p < 0.01, *** p > 0.001

This study had three main limitations. First, the focus was exclusively on breakoffs and given the priming of respondents during the invitation and landing page with regard to the length of the questionnaire, it could be that the impact of the questionnaire length is seen in the response rates rather than

Table 4. Cox Regression of Question Break-off: Interaction with Survey Phase

	1	2	3	4
Cond. Incentive X Survey Phase=2	0.601			
	-1.81			
Cond. Incentive X Survey Phase=3	-0.151			
	(-0.45)			
Cond. Incentive X Survey Phase=4	0.068			
	-0.21			
Screen: Medium X Survey Phase=2		0.956		
		-0.81		
Screen: Medium X Survey Phase=3		1.578		
		-1.44		
Screen: Medium X Survey Phase=4		-0.588		
		(-0.41)		
Screen: Large X Survey Phase=2		0.209		
		-0.68		
Screen: Large X Survey Phase=3		0.171		
		-0.57		
Screen: Large X Survey Phase=4		0.262		
		-0.91		
Female X Survey Phase=2			0.049	
			-0.16	
Female X Survey Phase=3			-0.303	
			(-1.03)	
Female X Survey Phase=4			0.12	
			-0.43	
Rural X Survey Phase=2				0.451
				-1.47
Rural X Survey Phase=3				0.21
				-0.72
Rural X Survey Phase=4				0.407
				-1.44
Observations	3,378	3,378	3,378	3,378

t-statistics in parentheses

* p < 0.05, ** p < 0.01; *** p > 0.001

the breakoffs. Second, the study was limited to three countries. It is tempting to take these countries as typical countries within their regions of Europe, but sampling and fieldwork processes differ dramatically from country to country. Finally, the analysis used a number of covariates to try and infer why individuals breakoff. In hindsight, it could have proved useful to have added a very short follow-up questionnaire, potentially via computer-assisted telephone interviewing to try and code reasons for breaking-off.

Despite these limitations, we draw three main conclusions. First, the breakoff rates varied widely by device type used, with 85% of respondents who used a large device (i.e., laptop or desktop computer) completing the whole survey. The largest breakoff rates were on smartphones, but we found no evidence they

were located around particularly difficult items for smartphones such as grids and loops. The layout of these questions was simplified to enable smartphone use, but it was still expected that the repetitive nature of these sections would be more tiresome on a phone. From the evidence, the experience of those choosing to fill the survey out on a phone needs to be radically improved.

Second, the results varied substantially across countries. One of the unique aspects of this study was its cross-national yet comparable design, and it is striking that the breakoff rates were lowest in Croatia when compared with Germany and Portugal. Germany has higher internet penetration and better internet speeds than Croatia and yet both the response rate and completion rate in Croatia were better. There is no obvious reason for this beyond the suggestion that Croatian respondents considered the survey as a novelty when compared with German respondents who are more accustomed to online data collection. Similarly, comparisons of urban and rural residents revealed no statistical difference in breakoff. However, this does indicate that apprehensions about fielding online surveys in Eastern and Southern Europe may be misplaced.

Third, the breakoff rates also differed between educational groups which likely effects the general willingness to participate, and the various skills required to fill in such a questionnaire. It should be noted that elements of the GGP questionnaire are particularly demanding and use relatively technical language to describe family life and demographic issues. The simplification of language should be considered to improve the accessibility of the questionnaire for lower educated groups.

Overall, the results are positive with regard to the prospects of long surveys being fielded online, but survey practitioners should place intensive efforts on reducing the number of complex and repetitive loops within a survey, ensuring that the design is optimized for all devices including smartphones and that a mix of unconditional and conditional incentives be used to ensure that breakoffs are minimized and responses are maximized. Finally, with regard to both questionnaire length and the sensitive nature of the questionnaire, it was concluded that a high-quality landing page and introductory letter that addressed these issues in advance was invaluable. Demographic change is an important topic that many people can relate to and are motivated to provide responses to. Clearly stating that participation involves a lengthy and sensitive questionnaire helps establish transparency and trust while maintaining that motivation.

.....

Acknowledgements

This work was funded by the Horizon 2020 Research and Innovation Programme under grant agreement no. 739511 for the project Generations and Gender Programme: Evaluate, Plan, Initiate and the Bundesinstitut für Bevölkerungsforschung.

Submitted: April 11, 2023 EST, Accepted: June 16, 2023 EST

REFERENCES

- Chen, Zeming, Alexandru Cernat, and Natalie Shlomo. 2023. "Predicting Web Survey Breakoffs Using Machine Learning Models." *Social Science Computer Review* 41 (2): 573–91. <https://doi.org/10.1177/08944393221112000>.
- Couper, Mick P., and Gregg J. Peterson. 2017. "Why Do Web Surveys Take Longer on Smartphones?" *Social Science Computer Review* 35 (3): 357–77. <https://doi.org/10.1177/0894439316629932>.
- Galesic, Mirta, and Michael Bosnjak. 2009. "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." *Public Opinion Quarterly* 73 (2): 349–60. <https://doi.org/10.1093/poq/nfp031>.
- Gauthier, Anne, Susana Laia Farinha Cabaço, and Tom Emery. 2018. "Generations and Gender Survey Study Profile." *Longitudinal and Life Course Studies* 9 (4): 456–65. <https://doi.org/10.14301/lcs.v9i4.500>.
- Gauthier, Anne, and Tom Emery. 2014. "Generations and Gender Programme: A Research Infrastructure For Analyzing Relationships over the Life-Course." In *Facing the Future: European Research Infrastructures for the Humanities and Social Sciences*, 99–108. http://forscenter.ch/wp-content/uploads/2014/05/2014_06_04-FACING_THE_FUTURE.pdf#page=83.
- Lambert, Amber D., and Angie L. Miller. 2015. "Living with Smartphones: Does Completion Device Affect Survey Responses?" *Research in Higher Education* 56 (2): 166–77. <https://doi.org/10.1007/s11162-014-9354-7>.
- Lugtig, Peter, Vera Toepoel, Tom Emery, Susana Laia Farinha Cabaço, Martin Bujard, Robert Naderi, Almut Schumann, and Detlev Lück. 2022. "Can We Successfully Move a Cross-National Survey Online? Results from a Large Three-Country Experiment in the Gender and Generations Programme Survey." *SocArXiv*, February. <https://doi.org/10.31235/osf.io/mu8jy>.
- Mittereder, Felicitas, and Brady T West. 2022. "A Dynamic Survival Modeling Approach to the Prediction of Web Survey Breakoff." *Journal of Survey Statistics and Methodology* 10 (4): 945–78. <https://doi.org/10.1093/jssam/smab015>.
- Peytchev, Andy. 2009. "Survey Breakoff." *Public Opinion Quarterly* 73 (1): 74–97. <https://doi.org/10.1093/poq/nfp014>.
- Steinbrecher, Markus, Joss Rossmann, and Jan Eric Blumenstiel. 2015. "Why Do Respondents Break off Web Surveys and Does It Matter? Results from Four Follow-up Surveys." *International Journal of Public Opinion Research* 27 (2): 289–302. <https://doi.org/10.1093/ijpor/edu025>.
- Toepoel, Vera, and Peter Lugtig. 2014. "What Happens If You Offer a Mobile Option to Your Web Panel? Evidence from a Probability-Based Panel of Internet Users." *Social Science Computer Review* 32 (4): 544–60. <https://doi.org/10.1177/0894439313510482>.
- Wolf, Christof, Pablo Christmann, Tobias Gummer, Christian Schnaudt, and Sascha Verhoeven. 2021. "Conducting General Social Surveys as Self-Administered Mixed-Mode Surveys." *Public Opinion Quarterly* 85 (2): 623–48. <https://doi.org/10.1093/poq/nfab039>.