This is a translation of this article which appeared in German:

Burgard, T., Bosnjak, M. & Wedderhoff, N. (2020). Konditionierungseffekte in Panel-Untersuchungen. Systematische Übersichtsarbeit und Meta-Analyse am Beispiel sensitiver Fragen [Conditioning Effects in Panel Studies. Systematic Review and Meta-Analysis for Sensitive Items]. *Psychologische Rundschau*, *71* (2), pp. 89-95. https://doi.org/10.1026/0033-3042/a000479

Conditioning Effects in Panel Studies. Systematic Review and Meta-Analysis for Sensitive Items

Tanja Burgard[1], Michael Bosnjak[1,2] and Nadine Wedderhoff[2]

[1]Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID), Trier

[2]Universität Trier

Abstract: Panel data are indispensable for investigating causal relationships and answering longitudinal questions. However, it is controversial how the repeated survey of panel participants affects the quality of panel data. The expected learning effect of repeated participation is called panel conditioning and can have both positive and negative consequences for the validity of panel data. Sensitive items in particular are expected to have an impact on the social desirability of the information provided. The available evidence on conditioning effects for sensitive questions suggests different effects depending on the type of question and has so far only been processed in the form of narrative reviews. In the present meta-analysis, conditioning effects are examined on the basis of the available experimental evidence (154 effect strengths from 19 reports), depending on the type of question, as well as the frequency and intervals between surveys (dosage effects). Standardized mean differences between experienced and fresh participants are analyzed by multi-level meta-regressions. The effects of previous surveys on the response behavior in subsequent waves are only minor. At present, it can therefore be assumed that the quality of panel data is not influenced to a relevant extent by conditioning effects. Limits of the present meta-analysis and relevant research gaps are discussed.

Keywords: conditioning effects, meta-analysis, panel data, sensitive items, social desirability

Relevance and Effects of Panel Conditioning

Panel data are indispensable for answering longitudinal questions and drawing causal conclusions. Because both their collection and the long-term maintenance of a participant pool are complex and expensive, open panel infrastructures exist in many disciplines that take care of this and are available to the research community. Examples are the GESIS Panel (Bosnjak et al., 2018), the Understanding America Study (Alattar, Rogofsky, & Messel, 2018), KAMOS (Cho, LoCascio, Lee, Jang, & Lee, 2017) and the LISS Panel (description of these infrastructures: Das, Kapteyn, & Bosnjak, 2018; Weiß et al., in press). Thus, resources are pooled and the objectivity of a survey is increased. A strengthening of such infrastructures is also called for in psychology (Bruder, Göritz, Reips, & Gebhard, 2014).

In order to accompany the establishment and use of such a service at the ZPID - Leibniz Institute, the question also arises which factors influence the quality of panel data. A known risk for the validity of panel data is panel mortality (Sobol, 1959). In many cases, it can be assumed that the departure of panel participants in later waves is related to relevant research variables and that the sample thus becomes less representative of the population over time. Another potentially detrimental effect is known as panel conditioning. This can have both positive and negative effects on data quality.

Basically, panel conditioning describes a learning effect in panel studies that is related to response behavior. It is assumed that experienced participants behave and respond differently than panelists who are participating for the first time. In contrast to panel mortality, however, panel conditioning is not a-priori associated with negative consequences for data quality. On the contrary, this effect may be based on different mechanisms that can have both positive and negative effects on the validity of the data, which are outlined here as examples along the relevant model of the survey process according to Tourangeau, Rips, and Rasinski (2000). According to this model, at least four steps are necessary to answer a question: Understanding the question, retrieving relevant information, processing and evaluating the information, and selecting the appropriate response option.

Already the understanding of a question could be influenced by previous survey experiences. An experienced participant perhaps understands both the question and the response options better, knows the rules of the interview and identifies the instructions and their objectives more precisely. This cognitive relief regarding the instructions could lead respondents to more often stating opinions instead of selecting categories such as "don't know", especially in complex attitudinal questions (Binswanger, Schunk, & Toepoel, 2013). Moreover, a survey can trigger reflection processes beyond the survey and lead to greater attention to and discussion of survey issues (Sturgis, Allum, & Brunton-Smith, 2009). This cognitive stimulation can change attitudes as well as knowledge (e.g. on demographic data such as income; Fisher, 2019) and have an effect on data in subsequent waves. In addition to cognitive stimulation related to survey topics, memory effects can also cause the reporting of more coherent attitudes (Bergmann & Barth, 2018).

The processing and assessment of the retrieved information in subsequent waves can also be influenced by previous surveys. In the case of a 'survey fatigue' that develops over time, test persons could try to give satisfactory answers while keeping the cognitive burden and time expenditure of the survey low. Krosnick (1991) calls this behavior satisficing. This also includes the so-called speeding (Schonlau & Toepoel, 2015), a process of answering questions as quickly as possible without actually looking for the most accurate answer.

The effort of the survey can furthermore be reduced by selecting a suitable answer. If respondents know the rules of the interview from previous surveys, they can avoid follow-up questions, for example, by answering filter questions negatively (Kreuter, McCulloch, Presser, & Tourangeau, 2011) or by reporting a smaller social network and thus shorten the survey (Silber et al., 2019).

Since conditioning effects, e.g. with regard to social desirability, are to be assumed to a greater extent for sensitive items, we concentrate on these in the present meta-analysis. An item can be classified as sensitive if it possesses at least one of the following three characteristics (Tourangeau, Rips, & Rasinski, 2000): 1. question demands a socially undesirable answer (e.g.: Do you regularly consume illegal drugs?), 2. question is perceived as intrusive and private (e.g.: How many sexual partners have you had in the last three years?), 3. question is particularly relevant in terms of data protection (e.g: Did you earn income in the last year which you did not report to the tax office?).

The evidence on panel conditioning effects for sensitive questions suggests different directions of impact in terms of data quality depending on the type of question. Due to the greater familiarity with the interview situation, respondents fear fewer consequences and respond more honestly to attitude questions. Thus, less socially desirable responses can be expected among experienced participants compared to new participants (e.g. Phillips & Clancy, 1972; Fowler, 1995; Nancarrow & Cartwright, 2007;

Binswanger, Schunk, & Toepoel, 2013). The first hypothesis thus assumes a reduction of social desirability bias in attitudinal questions:

> H1: Experienced panelists answer sensitive attitudinal questions less socially desirable than new panelists.

In the case of socially undesirable behavior, it is argued that their reporting triggers negative emotions, such as guilt, shame or fear, and thereby initiates a reflexive process that leads to the adaptation of responses towards social conformity in subsequent waves (Baumeister, Vohs, DeWall, & Zhang, 2007). Especially in studies on drug abuse among adolescents and young adults, evidence of this so-called recanting effect is found (Percy et al., 2005). Already reported drug abuse is denied in subsequent waves (Torche, Valenzuela, Warren, & Halpern-Manners, 2012). Similar effects have also been found in the context of other sensitive behavioral issues (Williams, Block, & Fitzsimons, 2006; Fitzsimons & Moore, 2008; Halpern-Manners, Warren, & Torche, 2014). This leads to the assumption of an increase in social desirability bias in sensitive behavioral questions with increasing survey frequency:

> H2: Experienced panelists are more likely to respond socially desirable than new panelists when it comes to sensitive behavioral questions.

A general assumption for conditioning effects, which applies to both attitudinal and behavioral items, is the existence of dosage effects. That is:

> H3: The more often the experimental group has already been interviewed, the stronger the conditioning effect, i.e. the difference between the standardized mean values of experienced and new participants.

> H4: The greater the time interval from the previous survey, the weaker the conditioning effect.

Methods

In order to be able to draw causal conclusions on conditioning effects, (quasi-)experimental studies of the response behavior in panel surveys are relevant for this meta-analysis. For this purpose, the responses of a previously interviewed experimental group and a control group not yet conditioned by interviewing must be compared at the same time. Both groups should have been surveyed about the same sensitive items, so that the corresponding behavior or attitudes can be compared.

Information was extracted from the relevant studies on three levels: 1. general information on the study report (author, year of publication), 2. description of the intervention (type of question, frequency of interview), 3. quantitative results of both groups (mean values, proportional values, test statistics, standard errors). A complete overview of all coding categories used is documented in Appendix 1.

A first literature search was conducted in December 2017 with the meta-search engine CLICsearch. Appendix 2 lists all databases included in CLICsearch. Besides "panel conditioning", 15 synonymous search terms were used (see Appendix 3). With the relevant articles identified after the first screening, a manual forward and backward search was additionally performed. This means that all cited and all referring literature entries were checked.

The calculated effect sizes are standardized mean differences. During coding and calculation, these are directed in such a way that positive values indicate that experienced panelists respond less socially desirable and negative values stand for a higher social desirability of the data in the test group. To test

the hypotheses accounting for the hierarchical data structure (several effect sizes were taken from the same study, for example), multilevel meta-regressions are used (Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013). A multilevel model also makes it possible to explain the distribution of variance in effect sizes by variables at different levels (Assink & Wibbelink, 2016). Appendix 4 documents and explains why a three-level model was chosen for the present meta-analysis.

All analyses were performed with the R-package metafor, version 2.0-0 (Viechtbauer, 2010).

Results

A total of 2 355 articles were examined to see whether they met the inclusion criteria. On the basis of the abstracts, 2 127 articles were excluded. The others were examined more closely and a further 209 items were not included in the meta-analysis due to the study design, the documentation of results or the non-sensitivity of the reported items. The 19 reports finally selected contain 85 samples and 154 effect sizes. The corresponding literature selection steps can be found in the form of a PRISMA flowchart (Moher, Liberati, Tetzlaff, & Altman, 2009) in Appendix 5.

Table 1 provides an overview of some characteristics of the 19 study reports included in the meta-analysis. In addition to author and year of publication, it also shows how the total number of samples and group comparisons are distributed among the publications. Most studies report results from a maximum of two samples. The effects of behavioral questions were examined more frequently (n=116) than those of attitudinal questions (n=38). Most studies were conducted in the USA. The mean standardized mean differences at the study report level are mostly close to zero, so the difference between control and experimental group is small. This finding speaks for no or only minor conditioning effects. However, there are some studies that suggest medium to strong effects, both in the direction of higher (negative SMD) and lower social desirability (positive SMD).

In order to quantify possible conditioning effects across all studies, meta-analyses were calculated with three levels of analysis (sample variance of effect sizes, variance between effect sizes, and variance between study reports).

((Table 1))

All estimated effects are very small (Ferguson, 2009) and not significant, as shown in Table 2. Across all 154 effect sizes, the mean SMD is positive. Experienced panelists thus answer less socially desirable than new participants across all studies and question types. In Hypothesis 1, exactly this was assumed for attitudinal questions. In fact, as expected, the sign of the estimated effect is positive for attitudinal questions, but the effect is not significant. Hypothesis 2, that expected a higher conformity with social norms for behavioral questions in the experimental group, is also not confirmed. All in all, it can be deduced that for behavioral and attitudinal questions no significant conditioning effects are to be expected in panel surveys. The differences between the groups and thus also the influence of previous surveys on data quality are very small.

((Table 2))

With hypotheses 3 and 4, dosage effects were established. On the one hand, conditioning effects should become stronger through more frequent questioning (hypothesis 3). On the other hand, it was assumed that the influence of previous interviews on the new measurement should decrease with increasing distance between waves (hypothesis 4). To test hypotheses 3 and 4, the absolute SMD is used as the effect size of interest. While in hypotheses 1 and 2 the direction of the differences between the groups was of interest, hypotheses 3 and 4 only refer to the strength of the effects, i.e. the absolute differences between the groups. The stronger the differences between the groups, the stronger the conditioning effect and vice versa. Thus, for hypothesis 3, a positive sign for the effect of the frequency of questioning can be expected. As the results in Table 3 show, the effect in the univariate random effects model is close to 0 and not significant. Even in the full model, which also takes into account the timing of the survey and the type of question, and in the model with interaction, the effect of the frequency of previous surveys remains negligible.

((Table 3))

Hypothesis 4 predicts a weakening of the conditioning effect with a greater distance between the survey waves. The results of the meta-regressions support this hypothesis. With greater distance between the waves, the difference between the control and experimental group is smaller, which shows the negative sign of the effect. This effect is also evident in the full model, in the case of attitudinal questions somewhat more pronounced than for questions on behavior.

Discussion and Conclusions

Overall, the meta-analysis on conditioning effects on sensitive attitudinal and behavioral questions allows the conclusion that previous surveys have only very little effect on the response behavior in subsequent waves. The available evidence does not suffice to make clear statements about the direction of the effects in relation to socially desirable response behavior. The assumption that the frequency of the survey strengthens conditioning effects could also not be confirmed. However, larger intervals between the surveys do indeed appear to have the presumed weakening effect on conditioning effects. In general, especially when evaluating dosage effects, the limited nature of the available database must be taken into account. There are hardly any planned experiments on conditioning effects in panels (Struminskaya, 2016). Most studies use panel refreshments to compare new participants with those already interviewed (Warren & Halpern-Manners, 2012). However, frequency and distances between waves cannot be specifically varied. For example, about half of the 154 group comparisons in the meta-analysis are based on experimental groups that were previously interviewed only once. Frequent intervals between panels are one week, one month and one year. There are hardly any gradations in between.

At present, no significant conditioning effects can be assumed in panel surveys. Panel-based surveys are and remain an important data source for psychology. However, the minor effects of panel conditioning found in the present analysis do not yet allow for final conclusions regarding the effects on the quality of panel data. Only the effects on sensitive attitudinal and behavioral items were considered. The data base on attitudinal items is already relatively limited, with 38 effect sizes from seven studies. Other types of questions, such as demographic data that could be considered confidential or non-sensitive filter

questions that could be answered incorrectly for strategic reasons in order to shorten the survey duration, were not examined here.

Panel conditioning is diverse and can have different causes and effects. In order to be able to make far-reaching statements about the quality and limitations of panel data, the various mechanisms must each be considered individually. This requires further meta-analyses, which should examine different types of questions and also different conditioning effects. Especially for the thorough investigation of dosage effects, which are particularly important for frequently surveyed populations such as participant pools of online panels, experimental primary studies are required in addition to meta-analyses. These allow a targeted variation of the survey dosage in order to close the gaps of previous research and to investigate timing effects more precisely.

References

Studies marked with * were used in the meta-analysis.

Alattar, L., Rogofsky, D. & Messel, M. (2018). An introduction to the understanding America study internet panel. *Social Security Bulletin*, *78* (2), 13–26.

Assink, M. & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*, 154–174.

* Axinn, W. G., Jennings, E. A. & Couper, M. P. (2015). Response of sensitive behaviors to frequent measurement. *Social Science Research*, *49*, 1–15.

* Barber, J. S., Gatny, H. H., Kusunoki, Y. & Schulz, P. (2016). Effects of intensive longitudinal data collection on pregnancy and contraceptive use. *International Journal of Social Research Methodology*, *19*, 205–222.

Baumeister, R., Vohs, K. D., DeWall, N. & Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review*, *8* (1), 1–20.

Bergmann, M. & Barth, A. (2018). What was I thinking? A theoretical framework for analysing panel conditioning in attitudes and (response) behaviour. *International journal of social research methodology*, *21*, 333–345.

Binswanger, J., Schunk, D. & Toepoel, V. (2013). Panel conditioning in difficult attitudinal questions. *Public Opinion Quarterly*, *77*.

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A. et al. (2018): Establishing an open probability-based mixed-mode panel of the general population in Germany: The GESIS Panel. *Social Science Computer Review*, *36* (1), 103– 115.

* Bridge, G. G. (1977). Interviewing Changes Attitudes – Sometimes. *Public Opinion Quarterly*, *41* (1), 56–64.

Bruder, M., Gö ritz, A. S., Reips, U.-D. & Gebhard, R. K. (2014). Ein national gefördertes Onlinelabor als Infrastruktur für die psychologische Forschung. *Psychologische Rundschau*, *65*, 75–85.

Cho, S. K., LoCascio, S. P., Lee, K.-O., Jang, D.-H. & Lee, J. M. (2017). Testing the Representativeness of a Multimode Survey in South Korea: Results from KAMOS. *Asian Journal for Public Opinion Research*, *4* (2), 73–87.

* Clinton, J.D. (2001). Panel Bias from Attrition and Conditioning. A Case Study of the Knowledge Networks Panel. *The American Association for Public Opinion Research (AAPOR)*, 56th Annual Conference.

* Coombs, L. C. (1973). Problems of Contamination in Panel Surveys: A Brief Report on an Independent Sample, Taiwan, 1970. *Studies in Family Planning*, *4*, 257.

* Das, M., Toepoel, V. & van Soest, A. (2011). Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys. *Sociological Methods & Research*, *40* (1), 32–56.

Das, M., Kapteyn, A. & Bosnjak, M. (2018). Open probability-based panel infrastructures. In D.L. Vannette & J. A. Krosnick (Eds.), *The Palgrave Handbook of Survey Research* (pp. 199–209). London, UK: Palgrave Macmillan.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. Professional Psychology: *Research and Practice*, *40*, 532–538.

Fisher, P. (2019). Does Repeated Measurement Improve Income Data Quality? *Oxford Bulletin of Economics and Statistics*, 0305–9049.

* Fitzsimons, G. J., John, D. S. N., Nunes, J. C. & Williams, P. (2007). License to Sin: The Liberating Role of Reporting Expectations. *Journal of Consumer Research*, *34* (1), 22–31.

Fitzsimons, G. J. & Moore, S. G. (2008). Should we ask our children about sex, drugs and rock & roll? Potentially harmful effects of asking questions about risky behaviors. *Journal of Consumer Psychology*, *18* (2), 82–95.

Fowler, F. J. (2006). *Improving survey questions: Design and evaluation*. Thousand Oaks, CA: Sage.

* Halpern-Manners, A., Warren, J. R. & Torche, F. (2014). Panel conditioning in a longitudinal study of illicit behaviors. *Public Opinion Quarterly*, *78*.

* Halpern-Manners, A., Warren, J. R. & Torche, F. (2017). Panel Conditioning in the General Social Survey. *Sociological Methods & Research*, *46* (1), 103–124.

* Kraut, R. E. & McConahay, J. B. (1974). How Being Interviewed Affects Voting: An Experiment. *Public Opinion Quarterly*, *37* (3), 398.

Kreuter, F., McCulloch, S., Presser, S. & Tourangeau, R. (2011). The Effects of Asking Filter Questions in Interleafed Versus Grouped Format. *Sociological Methods & Research*, *40* (1), 88–104.

Krosnick, J.A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, *5* (3), 213-236.

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: *The PRISMA Statement*. *PLoS Med 6* (7): e1000097.

*Murray, M., Swan, A. V., Kiryluk, S. & Clarke, G. C. (1988). The Hawthorne effect in the measurement of adolescent smoking. *Journal of epidemiology and community health*, *42*, 304–306.

Nancarrow, C. & Cartwright, T. (2007). Online access panels and tracking research: the conditioning issue. *International Journal of Market Research*, *49*, 573–594.

Percy, A., McAlister, S., Higgins, K., McCrystal, P. & Thornton, M. (2005). Response consistency in young adolescents' drug use self-reports: a recanting rate analysis. *Addiction*, *100*, 189– 196.

Phillips, D. L. & Clancy, K. J. (1972). Some Effects of „Social Desirability" in Survey Studies. American *Journal of Sociology*, *77*, 921–940.

* Quick, A., Bohnke, J. R., Pickett, K. E. & Wright, J. (2017). Does involvement in a cohort study improve health and affect health inequalities? A natural experiment. *BMC Health Services Research*, 17:79.

Schonlau, M. & Toepoel, V. (2015). Straightlining in Web survey panels over time. *Survey Research Methods*, *9* (2), 125–137.

Silber, H., Schröder, J., Struminskaya, B., Stocké, V. & Bosnjak, M. (2019). Does panel conditioning affect data quality in ego-centered social network questions? *Social Networks*, *56*, 45–54.

Sobol, M. G. (1959). Panel Mortality and Panel Bias. *Journal of the American Statistical Association*, *54* (285), 52–68.

* Song, Y. (2017). Rotation group bias in current smoking prevalence estimates using TUS-CPS. *Survey Research Methods*, *11*, 383–404.

* Struminskaya, B. (2016). Respondent Conditioning in Online Panel Surveys: Results of Two Field Experiments. *Social Science Computer Review*, *34* (1), 95–115.

Sturgis, P., Allum, N. & Brunton-Smith, I. (2009). Attitudes Over Time: The Psychology of Panel Conditioning. In: Lynn, P. (Ed.). *Methodology of longitudinal surveys* (pp. 113–126). Chichester, UK: Wiley.

* Toepoel, V., Das, M. & VanSoest, A. (2008). Effects of design in web surveys: Comparing trained and fresh respondents. *The Public Opinion Quarterly*, *72*, 985–1007.

* Torche, F., Valenzuela, E., Warren, J. R. & Halpern-Manners, A. (2012). Panel conditioning in a longitudinal study of adolescents' substance use: Evidence from an experiment. *Social Forces*, *90*, 891–918.

Tourangeau, R., Rips, L. J. & Rasinski, K. (Eds.). (2000). *The psychology of survey response*. New York, NY: Cambridge University Press.

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F. & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*, 576–594.

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

* Warren, J. R. & Halpern-Manners, A. (2012). Panel Conditioning in Longitudinal Social Science Surveys. *Sociological Methods and Research*, *41*, 491–534.

* Waterton, J. & Lievesley, D. (1989). Evidence of conditioning effects in the British social attitudes panel survey. In: D. Kasprzyk (Ed.). *Panel surveys* (pp. 319–339). New York, NY: Wiley.

Weiß, B., Das, M., Kapteyn, A., Bosnjak, M. & Schaurer, I. (2020). Open probability-based panels. *Wiley StatsRef: Statistics Reference Online*.

* Williams, P., Block, L. & Fitzsimons, G. (2006). Simply asking questions about health behaviors increases both healthy and unhealthy behaviors. *Social Influence*, *1* (2), 117–127.