GenPopWeb2: Adjustments for Mode Effects

Report: GenPopWeb2 meeting of experts (23rd September 2020)

Olga Maslovskaya[1], Lisa Calderwood[2], George Ploubidis[2], Gerry Nicolaas[3]

[1]Department of Social Statistics and Demography, School of Economic Social and Political Science, University of Southampton
[2]Centre for Longitudinal Studies (CLS), University College London (UCL)
[3] National Centre for Social Research (NatCen)

First version: 04 December 2020

Final version: 05 December 2023

This report summarises the survey methodology experts' discussion which took place on the 23rd of September 2020 at a closed event organised by GenPopWeb2 ESRC-funded project. The discussion addressed issues associated with **adjustments for mode effects**. The attendees at the event were Lisa Calderwood (CLS, UCL), Allison Callum (ONS), Paul Clarke (University of Sussex), David Hussey (NatCen), Charles Lound (ONS), Olga Maslovskaya (University of Southampton), Salah Merad (ONS), Kevin Pickering (Ipsos), George Ploubidis (CLS, UCL), Ria Sanderson (ONS), Barry Schouten (CBS Netherlands), Paul Smith (University of Southampton), and Peter WF Smith (University of Southampton).

Summary

Many social surveys have moved to mixed-mode designs, often motivated by potential cost-savings, but these designs can potentially increase risk of mode effects. In the absence of access to experimental design data, it is very important to address issues of endogeneity and potential mode effects in mixed-mode surveys by isolating mode effects on measurement from selection effects and by making necessary statistical adjustments. Currently there is no agreed best practice guideline and the methods used for isolation and statistical adjustment for mode effects vary. It is very rare for surveys to provide guidance to users on this issue. The choice of an approach is often subject specific or survey specific, or mode effects are often not considered at all by data users during the analysis. This practice increases the risk of reporting unreliable results. Statistical adjustments are not easy to implement as they are currently very technically demanding. Preventing mode effects is, of course, ideal but even with appropriate questionnaire design, prevention is not always possible or is often not fully successful. Moreover, the aim of preventing mode effects may conflict with trying to achieve the 'best' measurement in each mode. The experts identified the methods which are available for statistical adjustment and highlighted limitations of each method. It was identified that there is a great need for further methodological research in the area. The complete list of areas for methodological research can be found at the end of this report. Currently many data analysts who analyse mixed-mode surveys tend to ignore the problem of mode effects as there is no user-friendly approach to adjustment available for the researchers and data depositors rarely provide guidance on this. It was recommended that this practice should change, and improved guidance should be provided.

Introduction

Many social surveys have moved to mixed-mode designs during the covid-19 pre-pandemic period, often motivated by potential for cost-saving, but these designs can increase risk of mode effects. For some surveys covid-19 pandemic served as a catalyst to transitioning to online data collection and mixed-mode designs. Most surveys make this transition without randomised experimental tests or parallel runs to assess the impact of this change in data collection approach. In the absence of access to experimental design data, it is very important to address issues of endogeneity and potential mode effects in mixed-mode surveys by isolating mode effects on measurement from selection effects and by making necessary statistical adjustments. It is important also to understand the impact of mode on measurement to assess not only mode differences in measurement but also to understand which mode provides a better measurement for specific items. The approaches could be extended to mixed-device surveys too.

This area has been identified by [GenPopWeb2 network] as a high priority area of research. It is important to improve the understanding of statistical adjustments for mode and device effects and to formulate best practice guides in the area which would be beneficial for all researchers and data analysts using mixed-mode survey data. The report of the discussion at the meeting will be shared through our network of partners with the hope that it will reach wide academic and non-academic audiences.

Currently there is no agreed best practice guideline and the methods used for isolation and statistical adjustment for mode effects on measurement vary. It is very rare when surveys provide guidance to users on this issue. The choice of an approach is often subject specific or survey specific, or mode effects are not considered at all during the analysis and this practice increases the risk of reporting unreliable results.

Different disciplines developed various approaches of addressing the issues of endogeneity. For example, in econometrics instrumental variable approach is commonly used in the context of double-hurdle or Heckman selection models (Heckman, 1979). Instrumental variables are variables that are correlated with the endogenous variable but uncorrelated with the error term conditional on other covariates (Greene, 2012). Biologists use so called negative controls approach (Lipsitch et al., 2010). They identify “control” responses which are assumed not to be related to the “treatment”. Appropriate tests are then used to test for treatment effects. “Spurious” treatment effects imply selection, otherwise we can conclude that there is indeed a treatment effect (or in our case a mode effect). There are also sensitivity analysis approaches available such as the E-Value and other approaches (VanderWeele and Ding, 2017; Altonji et al. 2005; Oster 2019). Even if the mode effect does not actually reflect casual effect but it is just an association, users would still need to correct for it in applied analysis.

Among survey methodologists the approach to adjustments for mode effects is that adjustments are not easy to do, so the best way is to prevent mode effects which, of course, is ideal but it is recognised that prevention is not always possible or can be not fully successful. Moreover, the approach of designing surveys to prevent or minimise mode differences is also potentially in conflict with the aim of achieving optimal measurement in each mode. It is important to formulate advice on this for researchers and data users.

There is not much literature on adjusting for mode effects, aside from discussion of very sophisticated methods. Some of the main adjustment for mode effects approaches discussed in literature are: regression modelling adjustment approach (adjustments are computed by regressing survey responses on mode, demographics, and other relevant variables); multiple imputation adjustment approach (conceptualised as missing-data problem); re-interview method; approach which uses covariates to control for selection effects, and also alternative approach which uses covariates but explain measurement effects (Klausch et al., 2017a; Klausch et al., 2017b; Kolenikov & Kennedy, 2014; Schouten et al., 2013; Schouten et al., 2022; Vannieuwenhuyze et al., 2014).

This report will define mode effects, discuss how mode effects could be prevented and present existing methods for adjustments, it will formulate users’ needs, and will then identify gaps and future directions for methodological investigations.

What are mode effects? How do they arise?

Different response rates and cost implications are reasons for different modes being introduced in surveys and the use of mixed-mode designs. Cost is essential as some people will respond in cheaper mode and this will help in saving costs. However, some mixed-mode designs can increase the risk of mode effects (e.g., face-to-face and web). There are different definitions of mode effects. The responses to the same question asked in different modes may be systematically different and this difference is often called *mode effect*. The mode effects consist of selection effect which shows how the characteristics of respondents on observed variables vary according to what mode they choose to use for survey completion, and a measurement effect, which shows the differences to the responses to the substantive questions produced by the mode in which they are asked.

However, even with a single mode, mode effects will still be present but it will be difficult to identify them without a comparator group. The challenge for mixed-mode surveys is in combining and jointly analysing data collected in different modes. Of course, this depends on how mode effects are defined.

In mixed-mode surveys respondents often get to choose from different modes and certain groups of people would prefer web, phone or face-to-face modes. If specific characteristics of respondents are correlated with the chosen mode, then regression error term is correlated with the mode and the analysis is confounded and results of the analysis are unreliable. This issue is called *endogeneity*. However, there is no agreed best practice for handling endogeneity in mixed-mode surveys.

If the assumption is that there is a mode effect, it means that some people are not giving the “true” answer in certain modes, and we should then adjust for it. Different devices within mode might introduce device effects, so there are potential differences within online mode too. Even if a respondent can choose a mode, it would not get rid of mode effects due to various reasons such as the presence of the interviewer (social desirability bias) that may lead to measurement error in a certain mode such as face-to-face in comparison to self-administered modes.

Some researchers believe that it is ok to have mode effects as different modes are used to fully employ their advantages and mode effect is a price to pay. However, with this approach, adjustments should be considered.

When switching modes, there is discontinuity issue. It is important to be aware of mode effects when faced with discontinuity problem too.

Also, there is a mode effect not just on one variable but on multiple variables. It is also observed on the joint distribution of these variables so it can impact on the correlations and the higher order variance.

Different types of questions have larger mode effects, for example, attitudinal questions are prone to larger mode effects. This can be explained by the fact that there is no concrete reality underpinning the response (an opinion with the response scale where the respondent is themselves determining the meaning of each point on that scale). Usually factual variables do not have mode effects but Likert scales are prone to substantial mode effects. Also, self-reported variables have increased mode effects. NatCen produced a report on the risk of mode effects for different questions (D’Ardenne et al. 2017).

Mode effects can be due to selection if allocation to mode is not random. Also, causal inference is not just about isolating selection, it is also about isolating/dealing with confounding from unobserved sources. Mode effects may reflect both non-ignorable missingness and measurement error and it is important to address both aspects in adjustments.

If mode effects are observed, adjustments need to be done, even if it is just a random noise or manifestation of random error, it is still important to get rid of it.

There are different contexts in which mode effects are observed and they all need to be addressed differently: for example, sequential modes vs parallel modes. There is no single answer for all approaches, all modes and all surveys. For example, it is very difficult to adjust for mode in longitudinal surveys. It is important to consider all these different contexts and to develop different approaches appropriate for them.

If an estimate for the mode effect is available, it can then be used. For example, if an estimate of web mode effect is available from elsewhere, then web responses could be adjusted with a Bayesian approach.

Ideally, the impact of mode effects on applied research questions should be prevented or adjusted for and the guidance should be provided to data analysts on how to do it. But what is the best way of doing it? The following sections of the report will address the issues of prevention and adjustment of mode effects.

How can mode effects be prevented?

Questionnaire design

It is important to design survey questions which are adapted to mode to eliminate potential mode effect. Examples from Labour Market Survey (LMS) designed by the Office for National Statistics (ONS) suggested that large differences were found between modes on education questions. Also, in mixed-mode opinion surveys, large differences were found in question about cigarette smoking. It is very important to address design of questions to prevent mode effects. Unimode approach in which questionnaires look the same and ask the same questions in the same format across modes was used to minimise differences. However, there is a counter opinion suggesting that it is possible to use multiple modes to their advantage and to improve measurements but when there is an attempt to get rid of the mode differences then different modes are not really used to their full potential. If face-to-face questions were used directly in online mode with no change, mode effect would be created based on the desire not to break the series.

However, appropriate questionnaire design can be used to prevent or minimise mode effects but the decision should be made whether time series or minimisation of mode effects are more important.

Central Bureau of Statistics (CBS) Netherlands investigated the best mode(s) for specific questionnaires. They scored questionnaires using a lot of characteristics such as being a difficult language, recall that is needed to answer questions, and other characteristics (30 characteristics in total) and made profiles of these questionnaires based on what was known was affecting mode effects. This exploration is very useful and guides the preferences for certain modes. Their findings suggested that different modes might be more preferable for specific surveys.

What is known about adjustments for mode effects?

There are different approaches to adjustment for mode effects (Schouten et al., 2022). It is important to apply survey and/or context specific approach and check whether potential adjustment variables are available. Any adjustments will not get rid of all biases, but it is important to try to adjust out as much selection as possible and then what is left could be attributed to mode effects on measurement.

If an indicator of mode is available, it can be added as a control variable and it is possible then to estimate the raw mode effects on the outcome variable.

Alternatively, it is possible to separate data into two datasets and compare results of analysis. This difference would indicate mode effects. It could give a good indication if mode is affecting the analysis. If results are close enough, it could then be possible to combine both datasets together and to do the analysis but if they are not, then it is confirmed that the problem of mode effects is identified.

However, mode effect can be due to unobservable selection so due to causal inference problem or due to missing data issue and also can be explained by measurement error. It could be due to all these things or one of them. By splitting the data, it could be possible to get one right and one wrong answer or all wrong. As the mode is observed, this could be good news as it provides some opportunities for reasonable adjustments that could be user friendly.

There are absolute and relative approaches to adjustment for mode effects. In order to apply measurement model, concept of truth is needed and it is important to define what the truth is. However, it might be hard to do this.

If it is possible to estimate mode effects, they can then be incorporated. There is a measurement model which would need to be corrected by mode effects. For example, if the weight is 1 in one mode, then the weight is something else for a different mode. In this case relative differences are of interest, not absolute. If the concept of truth is available, absolute difference for each mode from the 'true value' can be obtained and necessary adjustments could be made. Different adjustments would need to be produced for different items within the same survey.

If the mode indicator is used as a dummy variable, then the assumption is that one mode is the truth (dummy variable's reference category). Alternatively, it is possible to relax the assumption that one mode is the truth but look at the relative differences between modes. This approach has limitations but is still better than doing nothing.

All methods will be adjusting to one "preferred" mode (usually face-to-face) or average value or mid-point between the different modes. Regression discussed below can be used to adjust for the non-preferred mode.

If a preference is not to choose the "true" mode, it is possible to report a measure of uncertainty instead or to use Total Survey Error (TSE) type approach and then conduct a pure variance estimation.

No approach would get rid of measurement error and selection. Therefore, it is important to decide whether point estimates or variance is more important when choosing the adjustment approach. The details of some adjustment approaches mentioned above will be discussed below.

Regression Approach

Regression approach to adjustment works in the following way: a model outcome variable is defined, auxiliary variables are added, mode effect is then added and estimated, if the rest of the model is correct so the estimate is indeed a mode effect. However, regression approach might be naïve and depends on the Directed Acyclic Graphs (DAG) that underlies the adjustment. For example, if mode effect is added to correct for income, potentially biased corrections might be obtained (as missingness issue can be present too). One could get biased adjustment because some of the auxiliary variables will be in both models (selection into mode and mode effect models). If model is not telling enough, one could end up with an estimator which is purely based on the preferred mode. It is also important to mention that adjustments depend on size of sample. It is important to be careful when half sample is adjusted to the other or part of the sample to the other that the efficiency - the precision - is not just determined by the sample one adjusts to. If 10% of sample is in face-to-face mode and 90% on a cheaper mode, it is important to be careful that adjustment does not end up effectively having a sample size of 1/10 of the total size.

Instrumental Variables Approach

If it is not possible to adjust for the mode effects using variables within the survey, this suggests that there are unobserved confounding factors driving who chooses which mode and this can be addressed with help of instrumental variables. They were widely used in one discipline but have not really transferred to other disciplines within last 10-15 years as it is just too difficult to choose instrumental variables. Strong instrumental variable should be predictive of the choice of mode but has nothing to do with the survey variables. In this approach it is important to make assumptions about unobservables.

Approaches from different disciplines are similar but language might be different. However, as mentioned earlier, instrumental variable approach is controversial as it is difficult to choose strong instrumental variables. It may be predictive of mode but it should have nothing to do with survey variables, no direct effect on responses to the surveys. The main problem with instrumental variables is that it can cause bias but equally can work well if valid.

Understanding Society and National Child Development Study (NCDS) age 55 both provide good examples of successful use of allocation to mode (initial randomisation) as instrumental variable. CBS Netherlands also has a good example of using different devices for survey response as successful instrumental variable (if respondents received an invitation in the form of a letter, they had to use a laptop for survey completion, if in the form of an email, they had to use a mobile device). However, it was concluded that region is a bad instrumental variable to use.

Auto-correlation

Auto-correlation in the series with repeated information for time series (discontinuity context) works well for switch of mode at a certain date.

Imputation Approaches

Mode effects issue can be viewed as missing data problem. “Hot deck” imputation approach (Andridge and Little, 2010) could be used to impute mode effects: if survey has two modes, so there are people who did web survey and their telephone responses could be set up as missing and for people who did telephone, their web responses should be set up as missing. These two groups of people can then be matched on the adjustment variables. It is then possible to look at the difference between two groups and that gives the mode effect for a specific person. It is possible to impute phone responses using a “donor” from the web responses based on matching characteristics. What happens if there is no exact donor? The problem is more obvious with non-parametric methods but it is hidden when using methods such as regression where value will always be imputed. Regression smooths this and always gives a prediction. Imputation would do the same. If observed confounding factors are suspected and those variables are available, the chosen approach depends on how robust one wants to be. Regression and parametric imputation will fill in and give estimates in a very efficient way but it would stretch data as there is a chance a “donor” value is used where it should not have been. To avoid this problem, the imputation could be restricted or a different link function could be used. Regression will find the effect of modes taking out the selection effect for everybody but that might not be the best thing to do. And this is the main difference between different methods. If it is impossible to find a “donor”, probably nonresponse mechanism is one of these nonignorable mechanisms.

It is possible to have many problems with matching, therefore imputation problem might become difficult. In practice it might not be possible to do it at all, unless it is a follow-up survey. Quality of adjustments depends on observables. Adjustment is only as good as variables and variables are only as good as the survey. However, it is impossible to get rid of all biases.

It is very important not to ignore survey nonresponse, as it is not known which mode would be preferred by the nonrespondents. Nonresponse would be nonignorable if it is impossible to find a "donor".

There is an issue that not much might be known about potential respondents if it is not a longitudinal context. So this approach would better work in a longitudinal context.

It is important not to overlook nonrespondents otherwise adjustments will be restricted only to respondents. It is important to make adjustment for their absence, i.e. for nonresponse. Should missing data handling only be restricted to the nonrespondents or whether it should be restricted only to those that responded to the survey and were allocated to a mode or self-selected to a mode? Depending on the nature of the survey the mode can be imputed trying to predict what would happen and what respondents would do depending on previous characteristics. In theory this could be achieved and this question is interesting in practice but not as straightforward as for other variables. This is a very important methodological question but a very difficult one to address too.

Fractional Imputation Approach

Fractional imputation was developed by researchers at the University of Iowa (Park et al., 2016; Kim and Yang, 2014; Yang and Kim, 2016). The approach worked well in two continuous variable context but not in multivariate, categorical or non-linear contexts, as it may be difficult to implement it. There was a structural model (regression) and a measurement error model, it was solved for continuous normal distribution but not categorical. It was then extended to incomplete bivariate ordinal response by She and Wu (2019). Park et al. (2016) assume one mode is a correct mode and an alternative mode is incorrect. They set up a joint model for effects of mode and a structural model. It is possible then to adjust that model. It is similar to an imputation model. Measurement effect model is adjusting within the multiple imputation framework for the mode effect which identifies in its simplest form by all the adjustment predictor variables that could be put in to explain the selection effect. Essentially it was driven by imputation design. It is taking responses from other mode and it is putting them in, matching them with people with the mode that one wants to impute. The approach is technically demanding as it has multiple imputation bases so there is lots of resampling and simulations going on. It works nicely in two variable approach they have but they did not go much beyond Y&X to the multivariate context so moving beyond that nonlinear categorical approach would be helpful. Theoretically they have got everything but implementing it in multivariate context where different variables would have various measurement issues across different predictors and outcomes would be a challenge. The model has two components: one is functional form of trying to translate information that one has to some relationships and how they affect each other and the other part is the data. The paper is very powerful on the functional side on how it could be used the kind of data if it is available but it does not provide solutions. Extension to two categorical variables is just a matter of time but what is overlooked is that the available information to understand even selection effects at best is usually weak. The authors have got combined instrumental variable approach at the end of the paper. If a good instrumental variable is not available, this approach is not going to be very useful for robustness.

Latent Variable Approach

Latent variable approach suggests that it might be possible in some situations to estimate what underlying true value is but it is important to have enough observations in both modes. This represents a real problem for cross-sectional context but might not be a complete problem in longitudinal studies where the same variable is measured multiple times, maybe even in multiple modes. This very fruitful approach is discussed in Sakshaug et al. (2021) in the context of Next Steps survey. It is possible with this approach to reach a point to say that this is not mode effect and then move to selection.

Re-interviewing Designs

In re-interviewing designs (Schouten et al., 2022) when respondents are re-approached under a second mode where relevant questions from the survey are repeated, respondents could be observed in two modes and this could be a great way forward to address issues of mode effects with confidence. However, the problem is that it is very rarely possible to change modes in reality and to go back to survey participants in practice.

Simulations

Pfeffermann (2019) conducted simulations and tried to apply a formal approach. He was correcting for selection effect. He proposed new estimator and used Bayes theorem. Estimation is the probability of being in one mode or another based on a vector of covariates. This might take away some of the selection but how can regression coefficients or descriptive statistics be adjusted? Pfeffermann (2019) conditioned on the mode and set of covariates and ran simulations. The approach is using the observed mode variable plus some observables from the survey in an appropriate way in a regression to adjust for the mode effect.

Triangulation

Triangulation approach can be used by combining instrumental variable approach, multivariate adjustments and sensitivity tests for unmeasured confounding factors.

All the approaches discussed above require a high level of technical expertise to implement and have various limitations. Due to complications linked to the adjustment for mode effects, the current practice advises to invest in design of the instruments and to focus on prevention and minimisation of mode effects rather than on adjustments where possible.

What should be done to help data users?

As mentioned earlier, currently data analysts tend to ignore the problem of mode effects as there is no user-friendly approach to adjustment for mode effects available.

Acknowledging mode effects is a responsible thing to do rather than pretending that the problem does not exist. It is better to do something than just to ignore the problem.

It is very important for data users that modes are flagged which is a routine. Also, it is very important to have information about overall survey quality in user manuals. User guides already contain information about weights and missing data. It is important to start including necessary information about mode effects to guides too. It is helpful to inform users of possible impact, which variables are most affected, and size of effects. Users then can do something and identify magnitude of effect if needed. If studies were conducted and impact is small, users should be informed about this too.

It is also important to report response patterns and description of the selection into mode pattern for all mixed-mode surveys in user guides. Users should be notified that mode effects might affect their conclusions.

Some suggestion should be made to users who do not have specialised skills but do data analysis on how to produce simple descriptive statistics and estimate regressions when there are different modes.

Statisticians should do the adjustments. Central analysis could be done without mode effects but then robustness checks where mode effects should be adjusted for should be conducted and the results might be included in the supplementary material sections. It is important to report the difference if any is found. Unfortunately, it is not common to see any mode adjustments in substantive papers and this should change.

Propensity scores of selecting a mode could be included together with data but depends on the aim of a specific project. In longitudinal context adjustments for mode effects might become more complex. It is important to run some form of simulations and some DAGs for specific family of models such as multilevel or survival models. This gets even more complicated in those contexts.

Imputed data should never be provided to users as various distortions could be introduced. Methods of multiple imputations should be available for users to decide how they want to implement them.

There are different groups of users which would benefit from different tools and methods. First, it should be identified what users want. Citizen users expect a single figure for a variable so need a method that is simple and can be explained well.

Policy-makers need to know if known imperfections will affect decisions. If the answer to this question is "yes", then they need to worry about changing the policy. And if the answer is "no", then the mode effect is not of concern. Their main concerns are to get the policy right. Will mode effects change policy decisions? If not, policy-makers do not need to worry.

Survey manages would be worried about making changes to data if the decision to adjust measurements is taken. Analysis should be adjusted but not data.

Users tend to dislike mode effects, therefore this view tends to dictate the survey question design (usually to be unimode) but this can also introduce mode effects to the data. It is important for users to know about design effects on mode effects.

What is unknown?

Future work: What methodological questions should be addressed?

The experts agreed that the following areas for methodological research are important:

1. Mode effect term is used in different ways, even to clarify this and to agree on definitions is an important starting point.
2. Survey methodologists need to come up with a framework for mechanisms as it is absent currently (response-nonresponse, selection into mode and then mode effect). It is hard to build estimation system if there is no framework. Important to set out process and phenomenon as it has not been described yet. It would be great to have SAGE booklet written on the topic.
3. There is a need to develop methods using Directed Acyclic Graphs (DAGs) as they can be useful to understand mechanisms and simulations. There is nothing for more complex methods

beyond regression, like multilevel models, survival modes currently available but it is important to extend the existing methods to other models.

4. Formal statistical simulations are needed to test new mode adjustment approaches and to address some nuances and subtleties of existing approaches such as regression approach, for example.
5. Various extensions of fractional imputation approach are needed. For example, extension to two categorical variables.
6. Latent class models should be developed in this context.
7. If auto-correlation is used in the series with repeated information for time series, it works well for switch of mode at a certain date but how would it work when respondents are choosing mode over a long period? Work is needed in this area.
8. It is important to answer the following question: What is a reasonable adjustment for specific families of models?
9. There is always descriptive statistics in each research paper which employs statistical analysis. Therefore, we need to address the following question: What impact adjustments would have on descriptive tables?
10. Statistical institutes often produce aggregate estimates. Methods that could be used in production for large number of surveys and for many variables within the same survey as well as estimating variances are needed. There is currently no method for estimating the variance available. It is very important for methodological work to get correct variance. The only context when variance can be obtained is time-series and discontinuity. There is a need in a guidance on how to estimate variances and need to have an established variance estimator.
11. It is important to extend investigation of patterns of selection into mode and to continue addressing the question which variables are more susceptible to mode effects which was started by NatCen (D'Ardenne et al, 2010). This work would be very useful for questionnaire design. If a quantitative measure of which variables are more or less susceptible to mode effect could be obtained, this would help to design the mode effects out. It is important to produce a risk index (three things to take into account: risky question, lots of selection which is difficult to account for and lots of measurement error) or score using PCA and pull out a main component. This score or index will tell the users how bad/good the situation is from the point of view of mode effects. If harmonised questions at risk of mode effect could be available, this would be a very useful catalogue. It is also important to understand harmonisation and mode and interplay between them.

Call for Action

It is important to focus on adjustments and to continue work in this area focusing on topics mentioned above. It is also important to produce a document for methodologists and also to organise an event for data users, analysts and PIs and other stakeholder groups probably separately to discuss the needs for different groups of users in the area of adjustments for mode effects.

References

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of political economy*, *113*(1), 151-184.

Andridge, R. R., & Little, R. J. (2010). A review of hot deck imputation for survey non-response. *International statistical review*, *78*(1), 40-64.

D'Ardenne, J., Collins, D., Gray, M., Jessop, C., & Pilley, S. (2017). Assessing the risk of mode effects: Review of proposed survey questions for waves 7-10 of Understanding Society. Understanding Society Working Paper Series, No.2017-04, April 017.

Greene, W.H. (2012) *Econometric analysis* (chapter 8), 7th Edition. Pearson.

Heckman, J. (1979) Sample selection bias as a specification error. *Econometrica*, 47, 153–61.

Kim, J. K., & Yang, S. (2014) Fractional hot deck imputation for robust inference under item nonresponse in survey sampling. *Survey Methodology*, *40*(2), 211-230.

Klausch, T., Schouten, B., Buelens, B., & Van Den Brakel, J. (2017a). Adjusting measurement bias in sequential mixed-mode surveys using re-interview data. *Journal of Survey Statistics and Methodology*, *5*(4), 409-432.

Klausch, T., Schouten, B., & Hox, J. J. (2017b) Evaluating bias of sequential mixed-mode designs against benchmark surveys. *Sociological Methods & Research*, *46*(3), 456-489.

Kolenikov, S. and Kennedy,C. (2014). Evaluating Three Approaches to Statistically Adjust for Mode Effects, *Journal of Survey Statistics and Methodology*, 2 (2), 126–158.

Lipsitch M., Tchetgen T.E. and Cohen T. (2010) Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology,* 21, 383–88.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187-204.

Park, S., Kim, J. K., & Park, S. (2016). An imputation approach for handling mixed-mode surveys. *The Annals of Applied Statistics*, *10*(2), 1063-1085.

Pfeffermann, D. (2019) Benefits and Issues in the Use of Internet-Based Surveys Experience from Israel. Presentation at the conference on the future of online data collection in social surveys, University of Southampton. https://www.ncrm.ac.uk/research/datacollection/Pfeffermann_SLIDES%20SOUTHAMPTON%20JUNE%202019_final.pdf

Sakshaug JW, Cernat, A, Silverwood S, Caldwerood L, Ploubidis GB (2021). Measurement Equivalence in Sequential Mixed-Mode Surveys. Survey Research Methods (In Press).

Schouten, B., van den Brakel, J., Buelens, B., van der Laan, J., & Klausch, T. (2013). Disentangling mode-specific selection and measurement bias in social surveys. *Social Science Research*, *42*(6), 1555-1570.

Schouten, B., van den Brakel, J., Buelens, B., Giesen, D., Luiten, A. and Meertens, V. (2022). *Mixed-Mode Official Surveys: Design and Analysis*. CRC Press, Boca Raton.

She, X., & Wu, C. (2019) Fully efficient joint fractional imputation for incomplete bivariate ordinal responses. *Statistica Sinica*, *29*(1), 409-430.

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: introducing the E-value. *Annals of internal medicine*, *167*(4), 268-274.

Vannieuwenhuyze, J.T.A., Loosveldt, G and Molnberghs, G. (2014). Evaluating Mode Effects in Mixed-Mode Survey Data Using Covariate Adjustment Models, *Journal of Official Statistics* 30(1), 1-21.

Yang, S., & Kim, J. K. (2016) Fractional imputation in survey sampling: A comparative review. *Statistical Science*, *31*(3), 415-432.