Mode effect: generally, effects that are due to the use of one mode compared to another, or a combination of modes to a single mode, or to a different combination of the same or other modes.

Mode selection effects: errors of non-observation. Combination of coverage and nonresponse.

Mode-measurement effects: observation errors (same respondent giving different answers to the same questions in different modes). AKA measurement bias or pure mode effects.

Concurrent design: multiple simultaneous modes

Sequential designs : One mode first then reapproach nonrespondents

Ways to deal with mode effects: weighting, imputation, other data processing

Analyze differences in the final sample composition based on different modes across time, countries and survey types, providing practical evidence-based guidelines for the National Statistical Institutes

Goals:

1. Overview on methodologies for mode effect assessment and adjustment, discussing assumptions, advantages and disadvantages of various approaches
2. Evaluate suitability of selected statistical approachyes and methods for selection and measurement effects in mixed mode data collection – practical application and statistical analysis
3. General guidance and assistance

We can often only observe the joint mode effect of selection/measurement, apart from experimental designs.

**Literature review**

De Leeuw (2005) advantages and pitfalls of mixing modes

Voogt and Saris (2005) trade-off between improved selection and possibly hampered measurements in multi-mode surveys

Dillman and Christian (2005) recognize the issue of differential measurement effects between modes and suggest preventing this issue through the design of questionnaires that prevent this phenomenon from occurring.

Fricker et al (2005) conducted an experiment comparing web and telephone surveys.

Experimental designs to disentangle mode selection/mode measurement: parallel, independent surveys or re-interview studies. In observational studies, relying on socio-demographic covariates that explain the selection mechanism -> reweighting or sample matching.

Mode adjustments: correct for bias introduced by one or several modes, require presence of a definition or choice of reference mode/benchmark. Techniques: reweighting and calibration, imputation, and prediction. Requires separation of selection/measurement effects.

**MIMOD**

Surveyed statistical offices in ESS countries on data collection strategies, questionnaire design, use of smartphones and tablets, methods of dealing with mode effects and case management systems.

Takeaways:

1. 1/3 of the countries did not assess mode effect in social surveys, but those that did used
   1. Pre-testing/experiments with questionnaire design were most common
   2. Experiments with sensitive or core questions
   3. Conducting pilot surveys
   4. Comparing distributions in socio-demographic or target variables
   5. Comparing quality indicators
   6. Parallel runs of different data collection strategies
2. 2/3 did not adjust for mode effects, but those that did predominately did so by weighting. A few also applied calibration or correction adjustments.
3. Half report future plans for researching mode effect assessment and/or adjustment methods
   1. Most focus on assessment, less on adjustment
   2. Plans for assessments are often quite rigorous, with pilot studies, experimental designs or parallel execution of different strategies.

Separating selection from measurement effects are a prerequisite for mode effect assessment and adjustment, this is important for the future. Re-interviews for second mode for respondents who responded through a first mode (Klausch, 2018).

**Applications**

Three options for dealing with mode effects when data collection modes are combined

1. Prevented through questionnaire design (Dillman (2014))
2. Avoided through data collection design (Schouten, Peytchev & Wagner (2017))
3. Adjusted through estimation design (Klausch et al (2018))

*Cost-benefit analysis of re-interview designs for mode-specific measurement bias*

Re-interview designs reapproach respondents to one mode by another mode, offering two measurements for part of the respondents in different modes, which can be used to estimate biases.

Reinterviews aren’t suited for concurrent mode designs where the respondent chooses the mode

Useful example: sequential web-telephone, web respondents reapproached by telephone at the same time as the telephone follow-up to web non-respondents

Assumptions: resp. unaffected by first mode contact and interview, non-response in second contact is unrelated to difference in measurement between the two modes.

1. Direct option: estimates mode-specific measurement bias
2. Indirect option : estimates mode-specific selection bias, then deduces mode specific measurement bias by subtracting the selection bias from the total mode bias

Direct: Two answers to a repeated question are compared, measurement differences are modelled and the estimated measurement bias model is applied to predict answers of those not in the reinterview, i.e. the nonrespondents to the first mode and to the re-interview.

Indirect: Response to the first mode is calibrated to the combined response to the re-interview mode and differences between the unadjusted and adjusted estimates are attributed to mode-specific selection bias. Estimated selection bias is subtracted from total bias to arrive at an estimate for the mode-specific measurement bias.

Both assume a benchmark design for bias estimation and adjustment

Schouten et al (2013) assumes mixed-mode design is selection benchmark, but one of the modes is the measurement benchmark.

Time-varying vs. constant is important.

Second condition is quality criterion adopted by main stakeholders: loss of precision from measurement bias adjustment. Stakeholders view the MSE of the final estimate as the criterion to judge quality (i.e. qilling to weigh a reduced precision against a reduced bias), and that the stakeholders constrain the precision to be the same after adjustment for measurement biases. Variance constraint setting implies the re-interview costs more.

*Health survey*

Estimates per mode:

*Table 2: Selected survey outcome variables with estimates per mode. Also provided is the estimated/anticipated reliability (correlation between repeated measurements).*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Survey* | *Estimate* | *Estimate* | *Correlation* |
| Unemployment rate | LFS 2014-2015 | 5.6 % | 6.7% | 0.5 |
| % good health | HS 2014 | 78.0% | 75.6% | 0.9 |
| % smoker | HS 2014 | 19.9% | 29.8% | 0.9 |
| % obese | HS 2014 | 12.1% | 13.9% | 0.9 |
| % visit to dentist | HS 2014 | 82.3% | 74.5% | 0.7 |

*Methods to asses and adjust mode effect on a social survey*

From Deliverable 3

Surveys based on mixed mode must be designed and carried out with constraints in mind for consistent and comparable estimates with analogues. (Changes in time series exclusively due to changes in observed phenomenon)

Application to ISTAT “Multipurpose Survey on Households – Aspects of daily life – 2017”. Mixed mode approach web + PAPI sequential. Parallel single mode PAPI to assess mode effect on two independent samples collected with different techniques.

Uses auxiliary info assumed to be mode insensitive from registers or survey. Goal Is evaluating impact of final estimates. Response and representativeness of two samples are evaluated through analysis of different nonresponse processes and representativeness indicators

Different operational steps:

1. Comparison between SM and MM sample
2. Evaluation of total mode effect in samples of resp. using web and PAPI in MM
3. Experiments to adjust for mode effects using MM data

Step 1) Identify set of relevant survey variables suspected of mode effect, look at difference in estimates. Then evaluate differences in response patterns in terms of magnitude of the bias that could explain differences in survey estimates between SM and MM. (Estimates are affected by total nonresponse differently in the two samples.) Evaluate bias introduced by total nonresponse with respect to a benchmark.

Step 2) Analysis of mode effect in MM sample – Propensity score. Equivalence of the measurements in the MM survey is analyzed based on the diagnostic method multi-group confirmatory factor analysis (MCFA). Correspondence tested for subjects responding using the web and PAPI techniques and mean level of latent factors useful for measuring the phenomenon with the techniques.

Step 3) Calibration of fixed proportions of web/PAPI responses has been applied to stabilize total measurement error over time. Counterfactual -> multiple imputation. Alternative estimates of the main parameters of the survey were obtained and compared with those produced by other adjustment methods.

Takeaways:

* MM reduces bias due to total nonresponse
* Total measurement error determined by different conflicting factors is hard to get at (response process/mode choice)
* Weighting can bring estimates in line in the adjustment phase
* Imputation can lead to very different results

**General discussion**

Deciding if/how to estimate/adjust for mode effects involves three key decisions

1. Quality criteria (e.g. MSE) against costs and assessment of whether/how beneficial mode effect adjustment is
2. Multi-dimensionality of a survey (key estimates and population parameters in need of evaluation)
3. Time perspective: repeated with constant effects

What mode is the benchmark?

Defining a methodological strategy has three components

1. A design
2. Auxiliary data/covariates (administrative data/frame/paradata)
3. Assumptions

Experimental designs can control for selection effects, but observational studies require covariates to explain selection mechanisms. Differences between mode groups are attributed to measurement differences, conditional on the covariates. (Can be validated on variables observed without error)

Three types of assumptions

1. About the explanation of the missing data mechanism due to mode selection
2. About the explanation of measurement differences due to modes
3. About the the absence of experimental influence on non-respondents

Steps to follow:

1. Identify the main quality and cost criteria
   * What benchmark is chosen for measurement?
   * Is it sufficient to consider accuracy (i.e. MSE) or also comparability in time and/or between subgroups?
   * What is the time horizon for which the mode design and budget are fixed and mode effects are estimated?
   * What are the key variables/population parameters of interest?
2. Decide whether mode effect estimation serves explanation only, design choice or adjustment
3. Identify available auxiliary data that is informative about
   * Mode selection
   * Mode measurement
4. Evaluate anticipated validity of assumptions for mode selection, mode measurement and absence of experimental influences
5. Decide whether an experimental design (such as re-interview or parallel run) is required and feasible to serve the purposes of the mode effect estimation;
6. Conduct experimental designs if deemed feasible and necessary

*Table 8. Analyses of total mode effect*

|  |  |  |  |
| --- | --- | --- | --- |
| **Objective of study: Assessing differences between estimates obtained based on data collected through different survey designs (single-mode and mixed-mode), in order to evaluate the total mode effect and the measurement equivalence** | | | |
| **Method** | | **Analysis** | ***Context* / Conditions** |
| Regression modelling approach to test whether design has a significant effect on the mean or  distribution of the item  (Martin and Lynn, 2011) | | Univariate analysis  of items to evaluate the impact on marginal distributions of mixed-mode design | * *Parallel independent surveys*   Appropriate statistical models and tests |
| Tests on differences in the estimates  (Martin and Lynn, 2011) | | Univariate analysis to highlight significant differences in the estimates calculated on the two sample designs | * *Parallel independent surveys*   Appropriate statistic tests for independent samples |
| Tests on indicators of *completeness* (item nonresponse)  Tests on indicators of *accuracy* (comparisons with external data)  (Jackle *et al*., 2010) | | Analysis on differences in the quality indicators | * *Parallel independent surveys*   Appropriate statistic tests |
| Multi-group confirmatory factor analysis  (Martin and Lynn, 2011; Hox *et al*., 2015) | | Analysis of the measurement equivalence when concepts are measured through more than one variable | * *Parallel independent surveys* * *Mixed mode survey design*   Identification of the latent structure of the phenomenon,  Control of selection effect |
| The proportional odds modelling technique (or parallel regression model, grouped continuous model)  (Jackle *et al*., 2010) | | Analysis to assess measurement equivalence of ordinal data on comparable samples | * *Parallel independent surveys* * *Mixed mode survey designs*   Control of selection effect  Validity of model assumption about covariates (covariates “shift” the distribution of responses proportionately across all categories) |
| Regression modelling approach whit one or more predictor variables and a binary indicator of single-mode and mixed-mode respondents  (Martin and Lynn, 2011) | | Multivariate analysis on estimates of the association between variables | * *Parallel independent surveys*   Appropriate statistical models and tests on significant interaction effects |
| **Objective of study: Analysing the response processes and evaluation of the bias caused by the total nonresponse (selection errors)** | | | |
| **Method** | **Analysis** | | **Context / Conditions** |
| Tests on the response rates respect to some characteristics of sample units  (Jackle *et al*., 2010) | Analysis on the response rates | | * *Parallel independent surveys* * *Single and mixed mode designs*   Appropriate statistic tests for independent samples |
| Summary statistic tests | Analysis of deviations from mode independence (absolute and relative selection error per benchmark variable) | | * *Parallel independent surveys* * *Comparison between single mode and mixed mode designs*   Appropriate statistic tests |
| R-indicator,  Conditional and Unconditional partial *R*-indicator  (Klausch *et al*., 2015; Schouten *et al*., 2011; Shlomo and Schouten, 2013; Schouten, *et al*., 2017) | Analysis of the representative response (absolute selection error for sets of benchmark variables) | | * *Parallel independent surveys* * *Single and mixed mode designs*   MAR assumption for Response model |
| Tests on the differences between benchmark variables (true value) and estimates  (Roberts and Vandenplas 2017) | Analysis on benchmark variables known for selected sample units | | * *Parallel independent surveys* * *Single and mixed mode designs*   Appropriate statistic tests |

|  |  |  |  |
| --- | --- | --- | --- |
| **Objective of study: Assessing mode effect - disentangling measurement and selection effects** | | | |
| **Method** | **Analysis** | **Conditions** | ***Context*** |
| Weighting   * Propensity score (PS) * Calibration * Post-stratification   (Vandenplas *et al*., 2016; Rosenbaum and Rubin, 1983; Vannieuwenhuyze, *et al*., 2014) | Analysis based on response model to control for respondent characteristics (comparable samples in MM) | MAR assumption  Mode-insensitive auxiliary variables  Balancing assumption in PS | * *Mixed mode survey designs (observational studies)* |
| Regression model  (Kolenikov and Kennedy, 2014) | Model analysis to estimate measurement and selection errors | Mode-insensitive auxiliary variables in the model to control selection effect | * *Mixed mode survey designs (observational studies)* |
| Other methods  -*double robust estimation* that combines an outcome regression with a propensity score model - *matching* | Model to estimate causal effect | Appropriate statistical models | * *Mixed mode survey designs (observational studies)* |
| Instrumental variable approach  (Vannieuwenhuyze *et al*., 2010) | Analysis based on benchmark single-mode design | Validity of comparability and representativity assumptions | * *Parallel independent surveys* |
| Re-interview  (Biemer, 2001) | Analysis based on re-interview data, administrative data and paradata.  The response of each mode is calibrated to the combined response of the re-interview and follow-up.  Measurement effect (ME) is estimated as remaining difference between modes.  Selection effect (SE) is estimated using mix of re-interview data, administrative data and paradata. | Re-interview does not affect measurement behavior of respondent.  Nonresponse to re-interview is unrelated to survey variables of interest given administrative data and paradata. | * *Re-interview of subset of mixed-mode respondents (experimental design with sequential mixed mode survey)* |

*Table 10. Approaches to adjust for mode effects*

|  |  |  |  |
| --- | --- | --- | --- |
| **Objective of study: Adjustment methods** | | |  |
| **Method** | **Data requirements** | **Assumptions** | **Advantages/Disadvantages** |
| Standard Covariate-based adjustment | * Sampling frame data * Paradata * Survey responses | Missing at random potential outcomes (MAR)  Exogeneity of auxiliary data | Too strong assumptions in many settings (-)  Adjustment on individual level possible (+) |
| Time-series stabilization/ mode calibration  (Buelens and van den Brakel, 2015, 2017) | Repeated cross-sectional / longitudinal survey | Independence of measurement and selection error  Time-stability of measurement error (ME) | Does not decompose (-)  Avoids ME estimation problem (+)  Strong assumption on mode contributions ( not fluctuate) (-) |
| Instrumental variable method  (Vannieuwenhuyze *et al*., 2010) | * Single-mode reference survey parallel to mixed-mode | Single-mode and mixed-mode survey have same selection bias (SB) | * Avoids MAR and exogeneity assumption (+) * Representativeness assumption usually implausible (-) * Not available for >2 modes |
| Re-interview method  (Klausch *et al*., 2017) | * Re-interview of subset of mixed-mode respondents | * Measurement equivalence | * More plausible MAR assumption (+) * MNAR estimators available (+) * Measurement equivalence traded off against true score time-stability (-) |

*Table 11 . Methods to adjust for mode effect*

|  |  |  |
| --- | --- | --- |
| **Objective of study: Adjusting selection/measurement effects in MM (observational studies)** | | |
| **Method** | **Aim** | **Conditions** |
| *Weighting*  - Propensity score  - Calibration  - Post-stratification  (Vandenplas *et al*., 2016; Rosenbaum and Rubin, 1983; Austin, 2011; Vannieuwenhuyze, *et al*., 2014) | To equate samples  To correct selection effect | Ignorability of selection mechanism (MAR)  Mode-insensitive auxiliary variables  Measurement error negligible |
| *Regression*  (Kolenikov and Kennedy, 2014) | To estimate measurement and selection effects  To correct measurement error | Appropriate statistical models |
| Other methods  -*double robust estimation* that combines an outcome regression with a propensity score model - *matching* | To estimate causal effect  To correct measurement error | Appropriate statistical models |
| *Multiple imputation* |  |  |
| 1.Multiple (standard) imputation | To predict counterfactual data (potential outcomes)  To correct measurement error | Choice of benchmark mode  MAR assumption |
| 2.Multiple imputation with response and selection models proposed by Suzer-Gurtekin *et al*. (2012) | Choice of benchmark mode  Sequential design and two modes  (Possibility – non-ignorability of selection mechanism) |
| 3.Fractional multiple imputation proposed by Park *et al*. (2016) | Sequential design and more than two modes  Possibility – non-ignorability of selection mechanism |