

## Evaluating Mode Effects in Mixed-Mode Survey Data Using Covariate Adjustment Models

Jorre T.A. Vannieuwenhuyze<sup>1</sup>, Geert Loosveldt<sup>2</sup>, and Geert Molenberghs<sup>3</sup>

The confounding of selection and measurement effects between different modes is a disadvantage of mixed-mode surveys. Solutions to this problem have been suggested in several studies. Most use adjusting covariates to control selection effects. Unfortunately, these covariates must meet strong assumptions, which are generally ignored. This article discusses these assumptions in greater detail and also provides an alternative model for solving the problem. This alternative uses adjusting covariates, explaining measurement effects instead of selection effects. The application of both models is illustrated by using data from a survey on opinions about surveys, which yields mode effects in line with expectations for the latter model, and mode effects contrary to expectations for the former model. However, the validity of these results depends entirely on the (ad hoc) covariates chosen. Research into better covariates might thus be a topic for future studies.

*Key words:* Selection effects; measurement effects; back-door model; front-door model; causal inference; opinion about surveys.

### 1. Introduction

Mixed-mode surveys are becoming increasingly popular for the collection of data from general populations (De Leeuw 2005, Voogt and Saris 2005, Dillman et al. 2009b, Vannieuwenhuyze and Loosveldt 2013). A *mixed-mode survey* is a survey in which data from different sample units is collected by different (sets of) data-collection modes. These include Computer-Assisted Personal Interviewing (CAPI), Computer-Assisted Telephone Interviewing (CATI), Postal Self-Administered Questionnaires (Postal SAQs), or Web Self-Administered Questionnaires (Web SAQs). The sample units can be defined either as individual sample members in cross-sectional data, or as time points within individual sample members in longitudinal surveys, so that each sample member is represented by different units.

Sample units can be selected for the data collection modes in four ways. First, in a sequential design, the modes are offered sequentially during a series of contact attempts. Second, in a concurrent design, all the modes are offered simultaneously during the first contact attempt and the sample members choose their preferred mode for responding.

<sup>1</sup> Institute for Social & Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, United Kingdom. Email: [jtavan@essex.ac.uk](mailto:jtavan@essex.ac.uk)

<sup>2</sup> Centre for Sociological Research, KU Leuven, Parkstraat 45, Leuven 3000, Belgium.  
Email: [geert.loosveldt@soc.kulueven.be](mailto:geert.loosveldt@soc.kulueven.be)

<sup>3</sup> I-BioStat, KU Leuven, Leuven, and Universiteit Hasselt, Diepenbeek, Belgium.  
Email: [geert.molenberghs@med.kulueven.be](mailto:geert.molenberghs@med.kulueven.be)

Third, in a comparative design, sample units are allocated to data-collection modes on the basis of some stratifying characteristics (for example, different countries use different modes in cross-national surveys, non-Internet households are approached by post instead of a web questionnaire, or different modes are used during different waves in a longitudinal survey). Fourth, in an allocative design, sample units are allocated to the data-collection modes in an experiment-wise random manner (however, each sample member can still choose whether or not to respond to the allocated mode).

Mixed-mode surveys are argued to have advantages over single-mode surveys because they may produce lower selection error, that is, the error introduced by only observing a small subset of the population instead of the entire population (De Leeuw 2005, Voogt and Saris 2005). First, a mixed-mode survey may reduce systematic selection error (e.g., nonresponse error or coverage error) compared to a single-mode survey, because certain members of the population might not be willing or able to respond to the mode used in the single-mode survey, but might respond to an alternative mode in the mixed-mode survey. In this case, the mixed-mode survey offers greater external validity than the single-mode survey. Second, a mixed-mode survey might reduce random selection error (e.g., sampling error) because some respondents may respond through a comparatively low-cost mode in a mixed-mode survey whereas the data-collection cost per unit would be higher in a single-mode survey. As a result, larger samples can be obtained within the same budget constraints. In this case, the mixed-mode survey offers greater external reliability than the single-mode survey.

As a consequence of the lower selection error, mixed-mode surveys provide, on average, samples that represent the population better compared to single-mode surveys, and thus parameter estimates that are closer to the population parameter or have smaller standard errors. However, it should be noted that the argument of lower selection error starts from the assumption that people's willingness to respond in a single-mode survey would persist in a mixed-mode survey that includes the same mode. This assumption might not hold in all situations because, for example, some studies observed lower response rates in a concurrent web and postal mixed-mode design compared to its postal only single-mode counterpart (Medway and Fulton 2012, Millar and Dillman 2011). Nevertheless, this assumption is further considered true throughout this article and we ignore situations where this assumption does not hold true.

Nevertheless, a necessary condition in order for mixed-mode surveys to obtain better representing samples is a *selection effect* between the modes, which means that sample units selected for the different modes differ on the variable of interest (Vannieuwenhuyze et al. 2012). Indeed, if selection effects are absent, then an alternative single-mode design will exist that uses the cheapest mode and provides data of equal external validity but higher external reliability. Evaluating the advantage of mixed-mode surveys thus primarily requires the estimation of selection effects. However, it must be noted that selection effects alone are not sufficient, as will be discussed in Section 5.

Further, evaluating selection effects in mixed-mode data is difficult because they are confounded with another type of *mode effect*: *measurement effects* (De Leeuw 2005, Voogt and Saris 2005, Dillman et al. 2009b, Weisberg 2005). Measurement effects are differences in measurement error accompanying the different data-collection modes (Voogt and Saris 2005, Weisberg 2005). Measurement effects thus occur when the answers given by the same respondents differ across the modes. As a consequence,

differences between the respondents in the alternate mode groups may either be due to differences in respondent characteristics (a selection effect) or to different measurement of responses (a measurement effect). Measurement effects therefore not only complicate the unbiased estimation of population parameters, but may also counteract the advantages of selection effects with regard to data quality.

The confounding of selection and measurement effects in mixed-mode data overlaps with a central theme of the causal inference literature (see, for example [Morgan and Winship 2009](#), [Pearl 2009](#), [Weisberg 2010](#)), which offers two distinct covariate adjustment models for disentangling selection and measurement effects and for obtaining unbiased estimates of population parameters ([Pearl 1995, 2009](#)). The first model requires covariates that capture selection effects, while the second model requires covariates that capture measurement effects. To date, both models have scarcely been theoretically discussed in literature relating to mixed-mode surveys. This article aims to fill the gap by providing a thorough theoretical discussion of both models, including the requirements, assumptions, advantages, and disadvantages.

The remainder of the article is structured as follows. Section 2 provides a brief discussion of the causal inference framework, including an overview of formal definitions of the mode effects. Section 3 provides a discussion of both covariate adjustment models and describes the required assumptions and estimation processes. Section 4 provides an illustration of the models using data from a survey about surveys. Section 5 finally concludes the article with a number of important suggestions for future research.

## 2. The Problem of Counterfactuals

For simplicity, this article is restricted to situations with only two data-collection modes, which we refer to as  $m_1$  and  $m_2$ . Further, the article is also restricted to the estimation of the population mean  $\mu$  on a variable of interest  $Y$ . Expansion into situations with more than two modes and more complex parameters can be derived straightforwardly from the following explanation, but may require more complex analysis frameworks.

The occurrence of measurement effects between modes means that the mode has a causal effect on the variable of interest and that respondents would have responded differently if different data-collection modes had been used. As a result, two potential outcomes are theoretically defined for each sample unit in which each potential outcome reflects the unit's outcome on variable  $Y$  if one particular mode had been used for data collection ([Rubin 1974](#), [Rosenbaum and Rubin 1983](#)). In the general causal inference literature, potential outcomes are traditionally represented on an aggregated level by two different variables, so that each unit is represented by one data line ([Holland 1986](#), [Rubin 1974](#)). In this article, by contrast, potential outcomes are represented on a disaggregated level by two different data lines per unit, because such disaggregated representation better allows for uniform definition of mode effects and model assumptions compared to the traditional aggregated representation.

The *full data* thus includes two data lines per sample unit, where each unit's first data line reflects the potential outcome when mode  $m_1$  was used, and the second data line reflects the potential outcome when mode  $m_2$  was used (see [Table 1](#)). The full data further requires definition of two additional variables. First, it requires a variable  $D$  that indicates

Table 1. The full data includes two data lines per unit, one observed and one counterfactual

Unit $U$	Mode group $G_\delta$	Mode of data collection $D$	Potential outcome $Y$	
1	$m_1$	$m_1$	$y_{1,m_1}$	= observed
1	$m_1$	$m_2$	$y_{1,m_2}$	= counterfactual
2	$m_1$	$m_1$	$y_{2,m_1}$	= observed
2	$m_1$	$m_2$	$y_{2,m_2}$	= counterfactual
3	$m_2$	$m_1$	$y_{3,m_1}$	= counterfactual
3	$m_2$	$m_2$	$y_{3,m_2}$	= observed
4	$m_2$	$m_1$	$y_{4,m_1}$	= counterfactual
4	$m_2$	$m_2$	$y_{4,m_2}$	= observed
$\vdots$	$\vdots$	$\vdots$	$\vdots$	

the distinction between the potential outcomes. This variable is further called the *mode of data collection* and takes the value  $m_1$  or  $m_2$ . Second, the full data requires a variable  $G_\delta$  that indicates the mode for which a unit is actually selected whenever this unit is a sample member of a mixed-mode survey with design  $\delta$ . This variable is further called the *mode group* and also takes value  $m_1$  when the respondent answers by mode  $m_1$ , and  $m_2$  when the respondent answers by mode  $m_2$ . It is important to stress that  $G_\delta$  is design specific. For example, some people may prefer mode  $m_2$  over  $m_1$  in a concurrent design, but would respond by mode  $m_1$  in a sequential design because they are unaware of the subsequent mode  $m_2$ .

Nonetheless, within *observed mixed-mode data*, only one data line is observed for each respondent because, by definition, all respondents in mode group  $m_1$  complete the survey by mode of data collection  $m_1$  instead of  $m_2$  and vice versa. Put differently, within mixed-mode surveys, data lines where  $G_\delta$  and  $D$  take different values are not observed (Table 1). For that reason, these data lines are called *counterfactual* (Galles and Pearl 1998, Greenland et al. 1999), but these counterfactuals are, nevertheless, important for the estimation of population means, selection effects, and measurement effects, as will be shown below.

The main objective of a survey is to obtain the best possible estimate of the population mean of the variable of interest. Ideally, the variable of interest is consistently measured over the entire population by one particular mode, which acts as a benchmark. For example, we can use mode  $m_1$  as the benchmark mode, because we believe mode  $m_1$  has a negligible measurement error while mode  $m_2$  is considered to be a distorting mode. As a consequence, the variable of interest is actually defined as  $(Y|D = m_1)$  and the population mean is defined as  $\mu_{m_1} = E(Y|D = m_1)$ , that is, the mean outcome when the values of all population members have been collected by mode  $m_1$ . The variable  $(Y|D = m_2)$ , in contrast, is a biased variable due to measurement error.

Using a mixed-mode design is believed to help obtain a sample that better represents the population, because some population members would not have responded if only one mode had been used, due to particular mode preferences or smaller possible sample sizes. The mixed-mode design thus would provide a better estimate of  $\mu_{m_1}$ . Nevertheless, unbiased estimation of the population mean  $\mu_{m_1}$  may still be difficult, because, by the law

of total expectation, it is the weighted sum of two conditional means where one mean requires counterfactual data for estimation:

$$\mu_{m_1} = \mu_{m_1 m_1} \tau_{m_1} + \mu_{m_1 m_2} \tau_{m_2}, \tag{1}$$

where  $\tau_g$  represents the unconditional probability  $P(G_\delta = g)$ , and  $\mu_{dg}$  represents the conditional mean  $E(Y|D = d, G_\delta = g)$ . The conditional mean  $\mu_{m_1 m_1}$  can be estimated from observed mixed-mode data, but the conditional mean  $\mu_{m_1 m_2}$  cannot be estimated without additional assumptions because it requires counterfactual data.

Furthermore, the population mean in (1) also clarifies why the estimation of the selection and the measurement effects is of primary interest for the evaluation of mixed-mode data quality. The conditional selection effect on the mean is the difference between the means of the people selected for modes  $m_1$  and  $m_2$  when all responses are measured by the same benchmark mode  $m_1$ :

$$S_{m_1}(\mu) = \mu_{m_1 m_1} - \mu_{m_1 m_2}. \tag{2}$$

If this selection effect is zero, then  $\mu_{m_1 m_2}$  would be equal to  $\mu_{m_1 m_1}$  and to the population mean  $\mu_{m_1}$ . In this situation, the population mean can be estimated straightforwardly by a single-mode design using mode  $m_1$ , which means that a mixed-mode design would be useless for increasing data quality compared to a single-mode design.

The conditional measurement effect on the mean is the difference between the means measured by the two different modes  $m_1$  and  $m_2$  for the same people who are selected for the distorting mode  $m_2$ :

$$M_{m_1}(\mu) = \mu_{m_2 m_2} - \mu_{m_1 m_2}. \tag{3}$$

If this measurement effect is zero, then  $\mu_{m_1 m_2}$  would be equal to  $\mu_{m_2 m_2}$  which can be estimated straightforwardly from the observed mixed-mode data. Put differently, a zero measurement effect would allow unbiased estimation of the population mean  $\mu_{m_1}$  with mixed-mode data, while a non-zero measurement effect would involve measurement bias on the population mean estimate.

Like the population mean, neither selection nor measurement effects can be estimated without additional assumptions because they require counterfactual data for the estimation of  $\mu_{m_1 m_2}$ . Indeed, the *overall mode effect*, which is the difference between the directly estimable conditional means of both modes, does not provide any information about the measurement and selection effects as it simply equals their difference, that is,

$$\mu_{m_1 m_1} - \mu_{m_2 m_2} = S_{m_1}(\mu) - M_{m_1}(\mu).$$

Put differently, it is not clear to what extent this difference is caused by a selection effect or a measurement effect. For that reason, selection effects and measurement effects are said to be *confounded* (Morgan and Winship 2009, Pearl 2009).

### 3. Analysis Models and Assumptions

The previous section made clear that the evaluation of mixed-mode data and the estimation of the population mean require estimation of  $\mu_{m_1 m_2}$ , which cannot be estimated directly because it requires counterfactual data. The task is to write down this mean in

terms of quantities that can be estimated by observed mixed-mode data, but that require analysis models with assumptions about relations between the variables. This section discusses two possible analysis models which include covariate adjustment.

Before continuing, note that selection and measurement effects are also defined by correlations between variables  $Y$ ,  $G_\delta$ , and  $D$  (see Figure 1; Pearl 1995, 2009). First,  $Y$  may relate to the mode group  $G_\delta$  due to unobserved common cause variables that simultaneously affect the variable of interest and the mode group for which a respondent is selected (as represented by curved bidirectional edges in Figure 1). The relationship between  $G_\delta$  and  $Y$  thus reflects a selection effect as it implies differences in respondent compositions between the mode groups. Second, by definition,  $Y$  is causally affected by the mode of data collection  $D$  (as represented by straight unidirectional edges in Figure 1), because the mode defines the measurement error in the response. The effect of  $D$  on  $Y$  thus denotes the measurement effect between the modes.

In the full dataset, where the responses of all respondents are observed in both modes  $m_1$  and  $m_2$ , there is no relationship between  $D$  and  $G_\delta$  (Figure 1a) because two data lines are theoretically defined for each respondent, one for each mode of data collection, irrespective of the actual mode group for which the respondent is selected in the mixed-mode survey. In the observed dataset, in contrast, the mode group  $G_\delta$  fully determines the mode of administration  $D$  for every respondent (as represented by the double-lined edge in Figure 1b) because all respondents in mode group  $m_1$  complete the survey by mode  $m_1$  instead of mode  $m_2$  and vice versa. As a result,  $G_\delta$  and  $D$  are equal and measurement and selection effects are completely confounded.

One could easily proceed by either assuming a zero selection or a zero measurement effect. A zero selection effect would mean that  $G_\delta$  and  $Y$  are unrelated (Figure 1c) and that respondents are completely randomly selected for the different modes. Such random selection overlaps with a proper experimental design and differences between both mode groups would be caused entirely by measurement effects. Nevertheless, a zero selection effect is not only unlikely but also unwanted as discussed in the previous section. A zero measurement effect, in turn, would mean that  $D$  and  $Y$  are unrelated (Figure 1d), that both modes come with equal measurement error, and that differences between both mode groups are entirely caused by selection effects. Nevertheless, like a zero selection effect, a zero measurement effect is very unlikely within mixed-mode surveys.

Instead of making improbable assumptions about zero selection and measurement effects, the literature about causal inference suggests the inclusion of adjusting covariates into the analysis model (Rosenbaum and Rubin 1983, Rubin 1974). Two types of covariates can be distinguished, where one type controls for selection effects and the other type controls for measurement effects (Pearl 1995, 2009). Both types are discussed in detail throughout the next subsections, which list the required model assumptions and show how both models allow the estimation of the counterfactual mean  $\mu_{m_1, m_2}$  if the assumptions hold true.

### 3.1. The Back-Door Model

The first analysis model with covariate adjustment involves the inclusion of a set of covariates  $B$ , where  $B$  is argued to explain the selection effect as a common cause of  $Y$  and

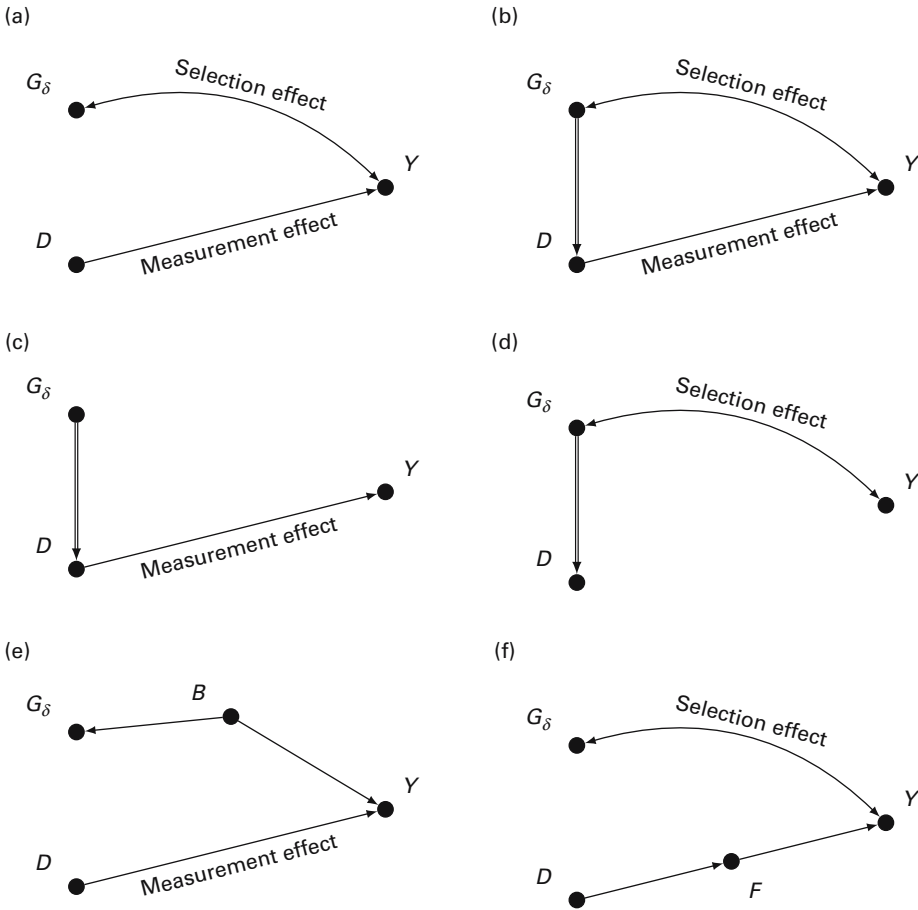


Fig. 1. Relationships between variables in mixed-mode data can be represented by causal graphs, where straight unidirectional edges represent direct causal effects and curved bidirectional edges represent correlations due to unobserved common causes (Pearl 1995, 2009). (a) In the full dataset, the mode group ( $G_\delta$ ) and mode of data collection ( $D$ ) are independent, and no confounding between measurement and selection effects occurs. (b) In a mixed-mode dataset, the mode group ( $G_\delta$ ) and mode of data collection ( $D$ ) are equal (double line), and measurement and selection effects are completely confounded. (c) The selection effect is zero when people are completely randomly selected for the different modes. The difference between the mode groups then equals the measurement effect. (d) The measurement effect is zero when all modes introduce equal measurement error. The difference between the mode groups then equals the selection effect. (e) Back-door covariates  $B$  allow for unbiased estimation of population means by blocking or explaining the selection effect. (f) Front-door covariates  $F$  allow for unbiased estimation of population means by blocking or explaining the measurement effect

$G_\delta$  (see Figure 1e). This model is called the back-door model by Pearl (1995, 2009), because it aims to capture ‘back-door’ correlations between the survey mode ( $G_\delta$ ) and the variable of interest ( $Y$ ) which arise from common cause variables.

Nevertheless, the back-door model starts from two assumptions (Pearl 2009, Morgan and Winship 2009). The first assumption is the *ignorable mode selection assumption* and requires that  $B$  fully captures the selection effect between the modes or that  $G_\delta$  and  $Y$  are

independent after controlling for  $B$  (as represented by the lack of an edge between  $G_\delta$  and  $Y$  in Figure 1e). If this assumption does not hold true, part of the selection effect is not captured and the confounding problem remains. The second assumption is the *mode-insensitivity assumption* and requires the absence of measurement effects on  $B$  or that  $D$  and  $B$  are independent (as represented by the lack of an edge between both variables in Figure 1e). If this assumption does not hold true, part of the measurement effect is channelled through  $B$  and the confounding problem again remains. It should, however, be noted that both assumptions cannot be empirically verified, as they refer to differences between observable and counterfactual outcomes.

If both the ignorable mode selection assumption and the mode-insensitivity assumption hold, it can be shown that the counterfactual mean  $\mu_{m_1 m_2}$  can be rewritten as an expression of quantities which can be estimated by observed data. For simplicity, let  $B$  be a discrete variable,  $\mu_{dgb}$  represent the conditional mean  $E(Y|D = d, G_\delta = g, B = b)$ , and  $\pi_{b|dg}$  represent the conditional probability  $P(B = b|D = d, G_\delta = g)$ . The following result emerges:

$$\begin{aligned}\mu_{m_1 m_2} &= \sum_b \mu_{m_1 m_2 b} \pi_{b|m_1 m_2} \\ &= \sum_b \mu_{m_1 m_2 b} \pi_{b|m_2 m_2}.\end{aligned}\tag{4}$$

The first step of (4) is an application of the law of total expectation. The second step follows from both assumptions. Indeed,  $\mu_{m_1 m_2 b} = \mu_{m_1 m_1 b}$  because  $Y \perp G_\delta|(D, B)$  by the ignorable mode selection assumption, and  $\pi_{b|m_1 m_2} = \pi_{b|m_2 m_2}$  because  $B \perp D|G_\delta$  by the mode-insensitivity assumption. As a result, implementing (4) into (1), (2), and (3) allows estimation of the population mean, the selection effect, and the measurement effect once an appropriate set of back-door variables is available:

$$\begin{aligned}\mu_{m_1} &= \sum_b \mu_{m_1 m_1 b} (\pi_{b|m_1 m_1} \tau_{m_1} + \pi_{b|m_2 m_2} \tau_{m_2}), \\ S_{m_1}(\mu) &= \sum_b \mu_{m_1 m_1 b} (\pi_{b|m_1 m_1} - \pi_{b|m_2 m_2}), \\ M_{m_1}(\mu) &= \sum_b \pi_{b|m_2 m_2} (\mu_{m_2 m_2 b} - \mu_{m_1 m_1 b}).\end{aligned}\tag{5}$$

Within the existing literature concerning causal inference, the back-door model is widely known due to the seminal work of Rubin (2005, 1991, 1978, 1974). Nevertheless, within Rubin's framework, the ignorable mode-selection assumption is formulated thoroughly, but the mode-insensitivity assumption is formulated less than clearly by the mere requirement that covariates must be collected at baseline (that is before treatment in an experimental study). As a result, within the existing literature concerning mixed-mode survey data, the back-door model has already been widely applied (see, for example Lutig et al. 2011, Heerwegh and Loosveldt 2011, Jäckle et al. 2010, Hayashi 2007, Fricker et al. 2005, Holbrook et al. 2003, Greenfield et al.



2000), but most of these studies use sociodemographic variables as back-door covariates. Such variables might easily be argued to be mode-insensitive, but they might not sufficiently explain why different people are selected for the different modes (Vannieuwenhuyze and Loosveldt 2013). Nonetheless, this issue is largely ignored within existing studies. Future studies might therefore focus on the search for better back-door covariates, such as paradata or survey questions asking for mode preferences (see, for example Olson et al. 2012).

### 3.2. The Front-Door Model

The second analysis model with covariate adjustment involves the inclusion of a set of variables  $F$ , where, in contrast to the back-door model,  $F$  is argued to explain the measurement effect as an intermediate variable between  $Y$  and  $D$  (see Figure 1f). This model is called the front-door model by Pearl (1995, 2009), because it aims to capture ‘front-door’ correlations between the survey mode and the variable of interest which arise from a direct causal effect of the survey mode ( $D$ ) on the variable of interest ( $Y$ ).

Like the back-door model, the front-door model also starts from two assumptions (Pearl 2009, Morgan and Winship 2009). The first is the *exhaustiveness assumption* and requires that  $F$  fully captures the measurement effects between the modes or that  $D$  and  $Y$  are independent after controlling for  $F$  (as represented by the lack of an edge between  $F$  and  $Y$  in Figure 1f). If this assumption does not hold true, part of the measurement effect is not captured and the confounding problem remains. The second assumption is the *isolation assumption* and requires the absence of selection effects on  $F$  or that  $G_\delta$  and  $F$  are independent (as represented by the lack of an edge between both variables in Figure 1f). If this assumption does not hold true, part of the selection effect is channelled through  $F$  and the confounding problem again remains. However, it should be noted that as with the back-door model, both assumptions cannot be empirically verified as they refer to differences between observable and counterfactual outcomes.

Similarly to the back-door model, if both the exhaustiveness assumption and the isolation assumption hold true, it can be shown that the counterfactual mean  $\mu_{m_1 m_2}$  can be rewritten as an expression of quantities which can be estimated by observed data. For simplicity, let  $F$  be a discrete variable,  $\mu_{dgf}$  represent the conditional mean  $E(Y|D = d, G_\delta = g, F = f)$ , and  $\pi_{f|dg}$  represent the conditional probability  $P(F = f|D = d, G_\delta = g)$ . The following result emerges:

$$\begin{aligned} \mu_{m_1 m_2} &= \sum_f \mu_{m_1 m_2 f} \pi_{f|m_1 m_2} \\ &= \sum_f \mu_{m_2 m_2 f} \pi_{f|m_1 m_1}. \end{aligned} \tag{6}$$

Once again, the first step of (6) is an application of the law of total expectation, while the second step follows from both assumptions. Indeed,  $\mu_{m_1 m_2 f} = \mu_{m_2 m_2 f}$  because  $Y \perp D|(G_\delta, F)$  by the exhaustiveness assumption, and  $\pi_{f|m_1 m_2} = \pi_{f|m_1 m_1}$  because  $F \perp G_\delta|D$  by the isolation assumption. As a result, implementing (6) into (1), (2), and (3) allows estimation of the population mean, the selection effect, and the measurement effect once

an appropriate set of front-door variables is available:

$$\begin{aligned}\mu_{m_1} &= \sum_f \pi_{f|m_1 m_1} (\mu_{m_1 m_1 f} \tau_{m_1} + \mu_{m_2 m_2 f} \tau_{m_2}), \\ S_{m_1}(\mu) &= \sum_f \pi_{f|m_1 m_1} (\mu_{m_1 m_1 f} - \mu_{m_2 m_2 f}), \\ M_{m_1}(\mu) &= \sum_f \mu_{m_2 m_2 f} (\pi_{f|m_2 m_2} - \pi_{f|m_1 m_1}).\end{aligned}\tag{7}$$

Even though the front-door model is analytically a mirror image of the back-door model, it is hardly mentioned in the literature on causal inference and we have found no mention to date in the literature on mixed-mode surveys. The front-door model requires variables that explain why people respond differently in the different modes. Therefore, front-door variables should try to measure, among other items, response burdens, satisficing, acquiescence, or social desirability. Potential front-door variables might be questions about, among others, survey pleasure or survey experiences (see, for example [Loosveldt and Storms 2008](#)), or variables including information about the number of item nonresponses or primacy and recency effects. For example, in Section 4, a question is used about whether the respondents found answering the survey a pleasant or unpleasant task. This variable provides results in line with expectations, even though it was selected ad hoc because the data was not collected with the idea of using the front-door model. The front-door model also therefore requires future research on the development and operationalisation of better front-door covariates.

## 4. An Illustration Using Data from a Survey About Surveys

### 4.1. Data Collection

The application of the back-door and front-door models will be illustrated by using them in connection with data from a survey concerning opinions about surveys, which was organised in 2004 in Flanders, Belgium, by the Survey Methodology Research Group of the Centre for Sociological Research, KU Leuven ([Storms and Loosveldt 2005](#)). The total sample consisted of 960 Flemish people aged between 18 and 80, sampled from the national register. A two-stage sampling procedure was used in which 48 communities were first selected with probability proportional to size and with replacement. Subsequently, 20 people were randomly drawn from each selected community. The clustering within communities is taken into account in the analyses and the data is weighted for differential nonresponse rates within the communities to preserve equal cluster sizes. Within-cluster nonresponse is further assumed to be ignorable.

A sequential mixed-mode design was used to collect the data ([Figure 2](#)). Each sample member was first contacted by post with an invitation to complete an enclosed paper questionnaire. If a sample member did not return the postal questionnaire, a first reminder was sent two weeks later and a second reminder accompanied by a new questionnaire was sent four weeks after the first reminder. The postal survey phase lasted two months in total. Sample members who did not return the paper questionnaire in due time were contacted by

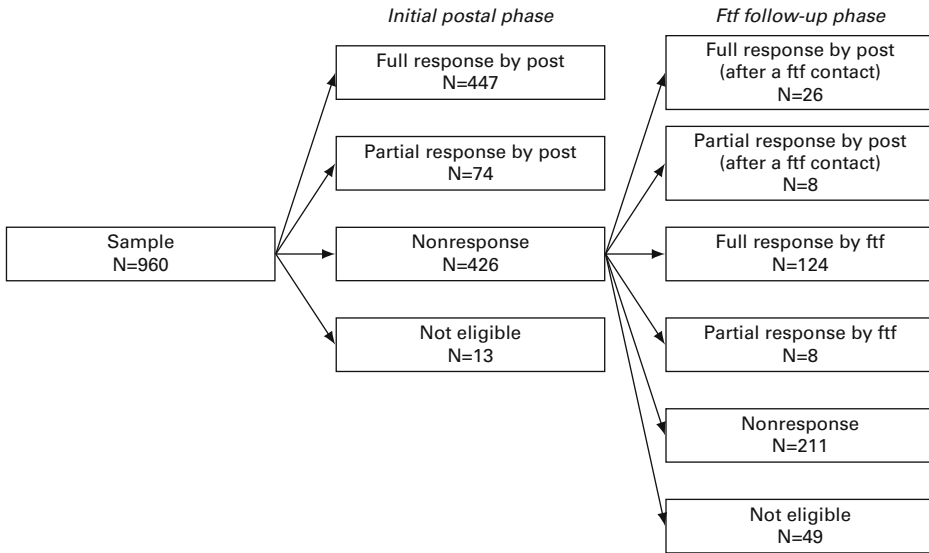


Fig. 2. The survey about surveys used a sequential mixed-mode design starting with a postal phase and ending with a face-to-face (ftf) follow-up

an interviewer at home to complete a face-to-face interview (i.e., CAPI). This face-to-face follow-up was not made known to the sample members during the initial postal phase.

For simplicity, the analyses will only include those respondents who responded to all the variables listed below. Only considering full responses, the initial postal phase reached a response rate (= full response/total sample – not eligible) of 47.20%, which the face-to-face follow-up increased to 63.04% (Figure 2). This response rate is relatively high for a general population survey.

## 4.2. Variables

### 4.2.1. Variables of Interest

Mode effects are analysed on the means of six items, each measuring a certain dimension of a short scale representing the respondents’ opinions about surveys (Loosveldt and Storms 2008). These items include statements about whether surveys are useful, whether surveys are a waste of people’s time, whether surveys stop people doing more important things, whether surveys are boring for respondents, whether the respondent likes surveys, and whether surveys are an invasion of privacy (Table 2). Respondents could indicate agreement or disagreement with these statements on a 5-point Likert scale ranging from ‘completely disagree’ to ‘completely agree’. In the postal questionnaire, these answer categories were listed horizontally in a table, but a ‘don’t know’/‘no opinion’ option was not provided. In the face-to-face interviews, the response categories were read out by the interviewer and presented vertically on a showcard, again excluding ‘don’t know’ and ‘no opinion’ options. For the analyses, all items were rescaled so that high values indicate positive opinions and low values indicate negative opinions.

Table 2. The survey about surveys includes six items/statements about surveys (Loosveldt and Storms 2008). Each respondent could indicate agreement or disagreement with each statement on a 5-point Likert scale (completely disagree, disagree, neither agree nor disagree, agree, completely agree)

Var.	Description
$Y_1$	'Surveys are useful ways of gathering information.'
$Y_2$	'Most surveys are a waste of people's time.'
$Y_3$	'Surveys stop people doing more important things.'
$Y_4$	'Surveys are boring for the persons who have to answer the question.'
$Y_5$	'I do not like participating in surveys.'
$Y_6$	'Surveys are an invasion of privacy.'

The particular topic of the survey might be very likely to cause selection effects and measurement effects on the means. First, there might be selection effects because nonrespondents to the postal questionnaire are likely to be more negative about surveys (Loosveldt and Storms 2008). The postal group data provide some evidence for this expectation: the later a postal questionnaire was returned, the lower the mean opinion score on all six opinion variables (table not included). Second, measurement effects are also expected, because respondents interviewed face-to-face will probably tend to report more positive opinions about surveys (Dillman et al. 2009a, Loosveldt and Storms 2008). Indeed, the mere presence of the interviewer may lead respondents to give socially desirable positive answers that do not reflect the respondents' real opinions.

#### 4.2.2. Back-Door Variables B

The back-door variables include a cross-classification of age and gender, educational level, ownership of a personal email address, activity status, and the number of adults (above 18 years of age), adolescents (between 12 and 18 years of age) and children (under 12 years of age) in the household. Age is divided into six categories, each spanning a period of ten years (18-27, 28-37, 38-47, 48-57, 58-67, and 68-80). The variable for educational level contains six categories: no qualification, primary school, lower secondary, upper secondary, college (non-university), or university. Activity status comprises eight categories: full-time employed, under 50% part-time employed, over 50% part-time employed, unemployed, retired, homemaker, disabled, and 'other'. The numbers of other people in the household also constitute different categories: 1, 2, 3, 4, and 5 or more adults, and 0, 1, and 2 or more adolescents or children.

These variables were chosen because they are very likely to be mode insensitive. Measurement effects are unlikely to occur between a face-to-face interview and a postal questionnaire on variables such as gender, age, the number of household members, or ownership of an email address. Firm evidence for the mode sensitivity of educational level and job-status variables is also lacking within existing literature, even though respondents might tend to overstate their educational attainment and describe themselves as employed when talking to an interviewer because they find these questions embarrassing (Lee and Renzetti 1990, Tourangeau and Yan 2007).

The central question is whether these back-door variables fully capture the selection effect on the variables of interest. Some insights can be provided by regression analysis of the back-door variables on the mode group and the variables of interest. These analyses

indicate significant associations between educational level and the mode group, but no significant associations between the back-door covariates and the variables of interest except for the number of adults and the question about privacy (item  $Y_6$ ) (tables not included). Although these associations therefore provide little evidence of possible selection effects, they nevertheless neither prove the absence of selection effects nor prove the capturing power of the back-door variables.

For the analyses, the set of back-door variables is transformed into one propensity score variable (Rosenbaum and Rubin 1983, Little 1986, Little and Rubin 2002). The respondents' propensity scores of responding via the postal questionnaire instead of the face-to-face interview are estimated by a maximum likelihood logistic regression model, using the mode group as the dependent variable and the back-door variables as independent variables. Subsequently, the estimated propensities are transformed into a grouped variable by coarsening the propensity scores into five values determined by using the 20th, 40th, 60th, and 80th percentiles as cut points. This coarsened propensity score variable is further used as the back-door variable  $B$ .

#### 4.2.3. Front-Door Variable F

As a front-door variable, a question is used which concerns the respondents' experiences during the survey. At the end of the questionnaire, the respondents were asked whether they found answering the questions a pleasant or unpleasant task. The respondents could select an answer from a 5-point Likert scale, comprising 'very pleasant', 'pleasant', 'neither pleasant nor unpleasant', 'unpleasant', and 'very unpleasant'. The format of this question in the postal questionnaire and the face-to-face interview was exactly the same as the opinion about survey items. Because relatively few respondents marked 'very pleasant', 'unpleasant', and 'very unpleasant', the variable was dichotomised ('very pleasant' and 'pleasant' versus 'neither pleasant nor unpleasant', 'unpleasant', and 'very unpleasant').

It is very likely that the mode of data collection has a direct causal effect on responses to the question about survey pleasure. The presence of an interviewer might intensify a feeling of discomfort because the respondent participated although he or she did not fully like the survey. Such a feeling of discomfort is resolved by adapting the reported attitude towards the actual behaviour, that is, by providing a socially desirable answer. Accordingly, the answers on survey pleasure from face-to-face respondents will be positive and consistent with eventual participation. Survey pleasure, in turn, probably has an effect on the reported opinion about surveys because people who report completing the survey as a pleasant task will tend to report more positive opinions about surveys in general.

The central question is whether this front-door variable fully captures the measurement effects on the variables of interest. Some insights can be provided by regression analysis of the mode group on the front-door variable and of the front-door variable on the variables of interest. There is a significant association between the mode group and survey pleasure (table not included). Moreover, even though the face-to-face mode includes more reluctant population members, the face-to-face respondents report a significantly higher pleasure compared to postal respondents. This observation may thus confirm the suggestion of cognitive dissonance. Likewise, the associations between survey pleasure and the opinion

items are always positive and also highly significant. These associations might thus provide some evidence of possible measurement effects. Nevertheless, these analyses neither prove the presence of measurement effects nor prove the capturing power of the front-door variable.

#### 4.3. Estimation Methods

The population means, selection effects, and measurement effects in (5) and (7) are functions of means and proportions which can directly be estimated from the data. The means are estimated by the SURVEYREG procedure in SAS, while the logits of the cumulative versions of the proportions are estimated by the SURVEYLOGISTIC procedure in SAS. These procedures take the clustered nature of the data into account as well as the random sample size of the population subgroups (or domains; see Cochran 1977). These procedures further also provide the covariance matrices of the estimates.

The resulting estimates of the SURVEYREG and SURVEYLOGISTIC procedures are maximum-likelihood estimates, which are known to be asymptotically normal with the mean equal to the population parameter. The Delta method, which uses first-order Taylor-expansions approximations (see Agresti 2002, Casella and Berger 2002, Lehmann 2001), can then be used to derive estimates for the population mean, the selection effects, and the measurement effects. In addition, the Delta method also provides approximate standard errors of the population means, selection effects, and measurement effects estimates, and proves that these estimates are also asymptotically normal.

#### 4.4. Results

The results show remarkable differences between the back-door and the front-door models with respect to the population mean estimates (Table 3). With the back-door model, the means are always larger when measured by a postal questionnaire ( $\mu_{\text{post}}$ ) compared to measurement by a face-to-face interview ( $\mu_{\text{fif}}$ ). With the front-door model, the opposite trend is revealed. In contrast to the back-door model estimates, the front-door model estimates are thus in line with the expectation that people represent themselves as more positive about surveys in front of an interviewer due to social desirability bias. Nevertheless, it must be noted that the differences between both modes are mostly small ( $<0.100$  on a 5-point scale).

With respect to selection effects, some differences are also found between the back-door and the front-door models. Taking the face-to-face interview as the benchmark mode (i.e.,  $S_{\text{fif}}(\mu)$ ), the back-door model does not yield large and significant selection effects. The front-door model, in contrast, does yield some significant negative effects. The negative signs of these significant selection effects are also in line with expectations, as they refer to more positive opinions of the postal respondents compared to the face-to-face respondents. The largest selection effect is found on the item about whether the respondent likes surveys (item  $Y_5$ ). This effect amounts up to  $-0.57$ , meaning that, on average, postal respondents rate their liking of survey participation 0.57 higher than face-to-face respondents on a 5-point scale.

Taking the postal questionnaire as the benchmark mode (i.e.,  $S_{\text{post}}(\mu)$ ), the back-door model yields one significant negative selection effect for the item about whether surveys

Table 3. The back-door and front-door models provide different estimates with respect to the population mean ( $\mu$ ), selection effects ( $S(\mu)$ ), and measurement effects ( $M(\mu)$ )

Effect	$\mu_{\text{fit}}$	$\mu_{\text{post}}$	$S_{\text{fit}}(\mu)$	$S_{\text{post}}(\mu)$	$M_{\text{fit}}(\mu)$	$M_{\text{post}}(\mu)$
Back-door model:						
$Y_1$	3.650*** 0.088	3.678*** 0.047	0.013 0.074	-0.065 0.040	-0.014 0.110	-0.038 0.104
$Y_2$	3.066*** 0.099	3.124*** 0.059	-0.022 0.092	-0.085 0.049	-0.027 0.127	-0.080 0.122
$Y_3$	3.320*** 0.094	3.360*** 0.054	-0.021 0.103	-0.097* 0.045	-0.053 0.141	-0.065 0.111
$Y_4$	2.991*** 0.094	3.058*** 0.052	-0.068 0.073	-0.065 0.043	-0.038 0.113	-0.095 0.112
$Y_5$	2.660*** 0.103	3.015*** 0.065	-0.155 0.086	-0.066 0.053	0.180 0.119	-0.401** 0.130
$Y_6$	3.431*** 0.082	3.501*** 0.048	0.089 0.088	-0.052 0.039	0.100 0.126	-0.063 0.095
Front-door model:						
$Y_1$	3.678*** 0.070	3.569*** 0.093	-0.119 0.094	0.073 0.122	-0.146*** 0.033	0.100 0.063
$Y_2$	3.098*** 0.089	3.034*** 0.106	-0.175 0.125	0.028 0.141	-0.181*** 0.042	0.033 0.073
$Y_3$	3.348*** 0.075	3.223*** 0.088	-0.153 0.102	0.076 0.116	-0.186*** 0.039	0.109 0.063
$Y_4$	3.023*** 0.074	2.928*** 0.094	-0.225* 0.108	0.099 0.125	-0.195*** 0.044	0.069 0.063
$Y_5$	2.746*** 0.095	2.654*** 0.099	-0.569*** 0.133	0.390** 0.133	-0.234*** 0.050	0.055 0.072
$Y_6$	3.482*** 0.071	3.361*** 0.078	-0.156 0.097	0.125 0.103	-0.146*** 0.034	0.114* 0.058

\*\*\*:  $p < .001$ , \*\*:  $p < .01$ , \*:  $p < .05$ , the  $p$ -values refer to two-sided tests of the null-hypothesis 'parameter=0'. For a description of the variables  $Y_1$  to  $Y_6$ , see Table 2.

stop people doing more important things ( $Y_3$ ), and the front-door model yields one significant positive selection effect for the item about whether the respondent likes surveys ( $Y_5$ ). A positive selection effect means that people selected for the postal questionnaire are more positive about surveys than people selected for the face-to-face survey when all data has been measured by the postal questionnaire. The positive front-door estimate for item  $Y_5$  is therefore again in line with expectations, because the face-to-face respondents were nonrespondents to the postal questionnaire. The negative back-door estimate for item  $Y_3$ , in contrast, is contrary to expectations.

With respect to measurement effects, the differences between the back-door and front-door models are even more striking. Taking the face-to-face interview as the benchmark mode (i.e.,  $M_{\text{fit}}(\mu)$ ), all back-door estimates are small and insignificant, but the front-door estimates are highly significant and negative. Moreover, all front-door estimates are negative and thus once again in line with expectations. Indeed, negative measurement effects mean that people responding through a postal questionnaire would report more positive opinions when surveyed in a face-to-face interview.

Taking the postal questionnaire as the benchmark mode (i.e.,  $M_{\text{post}}(\mu)$ ), the back-door model yields one significant negative selection effect for the item about whether the respondent likes surveys ( $Y_5$ ), and the front-door model yields one significant positive selection effect for the item about whether surveys are an invasion of privacy ( $Y_6$ ). Once again, the positive front-door estimate is in line with expectations and the negative back-door estimate is not. Indeed, positive measurement effects here mean that people responding in a face-to-face interview would report less positive opinions when surveyed using a postal questionnaire.

Last, the results also show striking differences between the measurement effects when the postal questionnaire and the face-to-face interview respectively are taken as the benchmark mode ( $M_{\text{post}}(\mu)$  and  $M_{\text{ffr}}(\mu)$ ). This difference may point to an interaction effect between measurement error and the mode group. People selected for the postal questionnaire seem to have larger measurement effects between both modes compared to people selected for the face-to-face interview.

#### 4.5. Discussion of the Illustration

To summarise, within the data from the survey examined, there is some evidence of selection effects between the modes, but the relevance of these selection effects may depend on the variable of interest, the analysis model, and on which mode is taken as the benchmark. Significant selection effects may point to a possible advantage of using mixed-mode data collection instead of single-mode data collection. Nevertheless, this advantage might not be guaranteed, because there is also evidence of measurement effects. These measurement effects may counteract the advantage provided by selection effects.

In general, large differences in estimates are observed between the back-door model and the front-door model. It should be emphasised that these differences are not caused by the models themselves, but by the variables that are selected as back-door and front-door covariates. It is very likely that the sociodemographic variables, which are used as back-door covariates, lack sufficient power to explain selection effects on the variables of interest. Further, it also remains unclear how much of the confounding of the selection and measurement effects is reduced by the front-door covariates. Nevertheless, because the front-door results were generally in line with expectations, the front-door covariates seem to perform better than the back-door covariates within this illustration.

## 5. General Discussion

The main aim of this article was to discuss the use of back-door and front-door models to disentangle selection and measurement effects and to estimate the population mean in mixed-mode survey data. Within relevant existing literature, studies concerning mode effect estimation chiefly use the back-door model, employing sociodemographic variables to explain selection effects. However, such sociodemographic variables probably do not meet the assumptions of the back-door model, which requires that the covariates both are mode insensitive and fully capture the selection effects. The front-door model, by contrast, remains largely unexplored within current literature regarding mixed-mode survey data. This model requires covariates which are assumed to both be insensitive to selection effects and fully capture the measurement effects between the modes.



This article widens the focus beyond the mere theoretical discussion of both the back-door and front-door models and aims to suggest a path for future research. Both the back-door and front-door models are theoretically sound ways of estimating population means, selection effects, and measurement effects, but the practical application of both models might offer challenges because mixed-mode data fit within the framework of so-called enriched data (Molenberghs et al. 2012). Enriched data, like, for example, incomplete data, censored time-to-event data, random-effects models, latent classes, latent variables, or mixture modelling, require strong and often empirically unverifiable assumptions. It is therefore imperative to carefully assemble the broadest possible evidence for the assumptions made in future studies on mixed-mode surveys. These future studies must, however, take the following points into account.

First, actual research on proper back-door as well as front-door covariates is all but nonexistent. Future research must start from other sources. A good source of candidates for back-door covariates might be questions about mode preferences (see, for example, Olson et al. 2012), whilst a good source of candidates for front-door covariates might be questions about survey pleasure or survey experiences (see, for example, Loosveldt and Storms 2008). Another possible source is paradata (see, for example, Kreuter et al. 2010) for both back-door and front-door covariates, but unfortunately the availability of such data might be very mode specific.

Second, the performance of back-door and front-door covariates largely depends on the survey design and the variable of interest. Mode effect estimates depend on the survey design through the mode group variable  $G_{\delta}$ , which is design specific. For example, the selection effects and measurement effects in a concurrent mixed-mode design might be different from those in a sequential design. As a consequence, different designs might require different back-door or front-door covariates. Further, mode effect estimates depend on the variable of interest because, for example, lower measurement effects are expected for factual questions than for sensitive questions about opinions. Once again, different kinds of variables of interest might require different back-door or front-door covariates.

Third, there is a need for research on the consequences of departures from the assumptions in both the back-door and the front-door models. Better knowledge of the relationship between the assumptions and mode effects estimation bias might not only help in selecting better covariates, but might also help in selecting optimal survey designs for particular survey topics.

Fourth, even though the back-door and front-door models are presented as two separate models, it should be noted that they can be integrated into the same analysis model. For example, the mode-insensitivity assumption of the back-door model requires the absence of measurement effects on the back-door covariates. Present measurement effects on back-door covariates may, however, be captured by a proper set of front-door covariates. These front-door covariates should not fully explain measurement effects on the variable of interest, but only on the back-door covariates. Likewise, back-door covariates can be used to capture present selection effects on front-door covariates and may guarantee the isolation assumption of the front-door model. The possibility of complex models provides additional opportunities for estimating mode effects and population means. Indeed, some back-door and front-door covariates might not perform well when used separately, but

may do a good job when combined into one analysis model. Nonetheless, it must also be kept in mind that more complex models may lead to estimation and identification problems.

Finally, it should be mentioned that in addition to the back-door and front-door models, a third model exists which also allows for estimation of mode effects. This model makes use of instrumental variables (Bowden and Turkington 1990, Angrist et al. 1996), but requires more complex survey designs and does not allow for estimating all conditional mode effects (Vannieuwenhuysen et al. 2012). Nevertheless, integration of the instrumental variable model, the back-door model, and the front-door model may also provide promising solutions.

Two remarks should be made in conclusion. First, this article describes the analysis of mode effects when only two modes are involved. Nevertheless, both the front-door and back-door models can also be applied when more than two modes are present. In that situation, researchers can use two strategies. In the first, they calculate the selection effects and the measurement effects between the benchmark mode and the other modes separately. In the second, they compare the distorting modes all together at once with the benchmark mode. This latter strategy is justified because the researcher may only be interested in measurement by the benchmark mode, while the separate contribution of the other modes to overall measurement bias is less important.

Second, it was stated in the introduction that the occurrence of selection effects is a primary condition for mixed-mode surveys to be advantageous, but their occurrence is nevertheless not a sufficient condition alone. Indeed, mixed-mode surveys involve higher fixed costs in terms of administration and organisation. An increase in these fixed costs might not be sufficiently compensated for by a decrease in the average cost per sample member through using a mixed-mode design. Especially for small samples, mixed-mode surveys might still not be advantageous over single-mode surveys even though selection effects occur. A cost-benefit analysis comparing mixed-mode and single-mode designs would be appropriate here. Such a cost-benefit analysis, however, first requires the estimation of mode effects and might thus provide a good topic for future studies.

## 6. References

- Agresti, A. (2002). *Categorical Data Analysis*. Hoboken, NJ: Wiley.
- Angrist, J.D., Imbens, G.W., and Rubin, D.B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, 91, 444–455. DOI: <http://www.dx.doi.org/10.1080/01621459.1996.10476902>
- Bowden, R.J. and Turkington, D.A. (1990). *Instrumental Variables*. Cambridge: Cambridge University Press.
- Casella, G. and Berger, R.L. (2002). *Statistical Inference* (2nd edition). Duxbury, CA: Pacific Grove.
- Cochran, W.G. (1977). *Sampling Techniques*. New York: Wiley.
- De Leeuw, E.D. (2005). To Mix or not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics*, 21, 233–255.
- Dillman, D.A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., and Messer, B.L. (2009a). Response Rate and Measurement Differences in Mixed-Mode Surveys Using

- Mail, Telephone, Interactive Voice Response (IVR) and the Internet. *Social Science Research*, 38, 1–18. DOI: <http://www.dx.doi.org/10.1016/j.ssresearch.2008.03.007>
- Dillman, D.A., Smyth, J.D., and Christian, L.M. (2009b). *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method* (3rd edition). Hoboken, NJ: Wiley.
- Fricker, S., Galesic, M., Tourangeau, R., and Yan, T. (2005). An Experimental Comparison of Web and Telephone Surveys. *Public Opinion Quarterly*, 69, 370–392. DOI: <http://www.dx.doi.org/10.1093/poq/nfi027>
- Galles, D. and Pearl, J. (1998). An Axiomatic Characterization of Causal Counterfactuals. *Foundations of Science*, 1, 151–182. DOI: <http://www.dx.doi.org/10.1023/A:1009602825894>
- Greenfield, T.K., Midanik, L.T., and Rogers, J.D. (2000). Effects of Telephone Versus Face-to-Face Interview Modes on Reports of Alcohol Consumption. *Addiction*, 95, 277–284. DOI: <http://www.dx.doi.org/10.1046/j.1360-0443.2000.95227714.x>
- Greenland, S., Pearl, J., and Robins, J.M. (1999). Causal Diagrams for Epidemiologic Research. *Epidemiology*, 10, 37–48.
- Hayashi, T. (2007). The Possibility of Mixed-Mode Surveys in Sociological Studies. *International Journal of Japanese sociology*, 16, 51–63. DOI: <http://www.dx.doi.org/10.1111/j.1475-6781.2007.00099.x>
- Heerwegh, D. and Loosveldt, G. (2011). Assessing Mode Effects in a National Crime Victimization Survey Using Structural Equation Models: Social Desirability Bias and Acquiescence. *Journal of Official Statistics*, 27, 49–63.
- Holbrook, A.L., Green, M.C., and Krosnick, J.A. (2003). Telephone Versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias. *Public Opinion Quarterly*, 67, 79–125. DOI: <http://www.dx.doi.org/10.1086/346010>
- Holland, P.W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81, 945–960. DOI: <http://www.dx.doi.org/10.1080/01621459.1986.10478354>
- Jäckle, A., Roberts, C., and Lynn, P. (2010). Assessing the Effect of Data Collection Mode on Measurement. *International Statistical Review*, 78, 3–20. DOI: <http://www.dx.doi.org/10.1111/j.1751-5823.2010.00102.x>
- Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R.M., and Raghunathan, T.E. (2010). Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Non-response: Examples from Multiple Surveys. *Journal of the Royal Statistical Society, Series A*, 173, 389–407. DOI: <http://www.dx.doi.org/10.1111/j.1467-985X.2009.00621.x>
- Lee, R.M. and Renzetti, C.M. (1990). The Problems of Researching Sensitive Topics: An Overview and Introduction. *American Behavioral Scientist*, 33, 510–528.
- Lehmann, E.L. (2001). *Elements of Large-Sample Theory*. New York: Springer.
- Little, R.J.A. (1986). Survey Nonresponse Adjustments for Estimates of Means. *International Statistical Review*, 54, 139–157.
- Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data* (2nd edition). London: Wiley.
- Loosveldt, G. and Storms, V. (2008). Measuring Public Opinions About Surveys. *International Journal of Public Opinion Research*, 20, 74–89. DOI: <http://www.dx.doi.org/10.1093/ijpor/edn006>

- Lugtig, P., Lensvelt-Mulders, G.J.L.M., Frerichs, R., and Greven, A. (2011). Estimating Nonresponse Bias and Mode Effects in a Mixed-Mode Survey. *International Journal of Market Research*, 53, 669–686.
- Medway, R.L. and Fulton, J. (2012). When More Gets You Less: A Meta-Analysis of the Effect of Concurrent Web Options on Mail Survey Response Rates. *Public Opinion Quarterly*, 76, 733–746. DOI: <http://www.dx.doi.org/10.1093/poq/nfs047>
- Millar, M.M. and Dillman, D.A. (2011). Improving Response to Web and Mixed-Mode Surveys. *Public Opinion Quarterly*, 75, 249–269. DOI: <http://www.dx.doi.org/10.1093/poq/nfr003>
- Molenberghs, G., Njeru Njagi, E., Kenward, M.G., and Verbeke, G. (2012). Enriched-Data Problems and Essential Non-Identifiability. *International Journal of Statistics in Medical Research*, 1, 16–44.
- Morgan, S.L. and Winship, C. (2009). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Analytical Methods for Social Research. New York: Cambridge University Press.
- Olson, K., Smyth, J.D., and Wood, H.M. (2012). Does Giving People their Preferred Survey Mode Actually Increase Survey Participation Rates? An Experimental Examination. *Public Opinion Quarterly*, 76, 611–635. DOI: <http://www.dx.doi.org/10.1093/poq/nfs024>
- Pearl, J. (1995). Causal Diagrams for Empirical Research. *Biometrika*, 82, 669–688. DOI: <http://www.dx.doi.org/10.1093/biomet/82.4.669>
- Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2nd edition). New York: Cambridge University Press.
- Rosenbaum, P.R. and Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70, 41–55. DOI: <http://www.dx.doi.org/10.1093/biomet/70.1.41>
- Rubin, D.B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66, 688–701. DOI: <http://www.dx.doi.org/10.1037/h0037350>
- Rubin, D.B. (1978). Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, 6, 34–58.
- Rubin, D.B. (1991). Practical Implications of Modes of Statistical Inference for Causal Effects and the Critical Role of the Assignment Mechanism. *Biometrics*, 47, 1213–1234.
- Rubin, D.B. (2005). Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association*, 100, 322–331. DOI: <http://www.dx.doi.org/10.1198/016214504000001880>
- Storms, V. and Loosveldt, G. (2005). *Procesevaluatie van het Veldwerk van een Mixed Mode Survey naar het Surveyklimaat in Vlaanderen*. Leuven: KUL, Centrum voor Sociologisch Onderzoek.
- Tourangeau, R. and Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin*, 133, 859–883.
- Vannieuwenhuyze, J.T.A. and Loosveldt, G. (2013). Evaluating Relative Mode-Effects in Mixed Mode Surveys: Three Methods to Disentangle Selection and Measurement

- Effects. *Sociological Methods and Research*, 42, 82–104. DOI: <http://www.dx.doi.org/10.1177/0049124112464868>
- Vannieuwenhuyze, J.T.A., Loosveldt, G., and Molenberghs, G. (2012). A Method to Evaluate Mode Effects on the Mean and Variance of a Continuous Variable in Mixed-Mode Surveys. *International Statistical Review*, 80, 306–322. DOI: <http://www.dx.doi.org/10.1111/j.1751-5823.2011.00167.x>
- Voogt, R.J. and Saris, W.E. (2005). Mixed Mode Designs: Finding the Balance Between Nonresponse Bias and Mode Effects. *Journal of Official Statistics*, 21, 367–387.
- Weisberg, H.F. (2005). *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: University of Chicago.
- Weisberg, H.F. (2010). [Bias and Causation: Models and Judgment for Valid Comparisons](#). Hoboken, NJ: Wiley.

Received April 2012

Revised April 2013

Accepted September 2013