ACCOUNTING FOR THE EFFECTS OF DATA COLLECTION METHOD. APPLICATION TO THE INTERNATIONAL TOBACCO CONTROL NETHERLANDS SURVEY

Accounting for the Effects of Data Collection Method

Application to the International Tobacco Control Netherlands Survey

Mary E. Thompson, Y. Celia Huang, Christian Boudreau, Geoffrey T. Fong, Bas van den Putte, Gera E. Nagelhout, Marc C. Willemsen

In **Population Volume 68, Issue 3, 2013**, pages 447 to 471

ISSN 0032-4663
ISBN 9782733201725

This document is the English version of:
Mary E. Thompson, Y. Celia Huang, Christian Boudreau, Geoffrey T. Fong, Bas van den Putte, Gera E. Nagelhout, Marc C. Willemsen, «Mesurer les effets de la méthode de collecte des données», Population 2013/3 (Vol. 68) , p. 447-471

Mary E. THOMPSON,* Y. Celia HUANG,* Christian BOUDREAU,*
Geoffrey T. FONG,* Bas VAN DEN PUTTE,**
Gera E. NAGELHOUT,*** Marc C. WILLEMSEN***

Accounting for the Effects of Data Collection Method

Application to the International Tobacco Control Netherlands Survey

Quantitative surveys employ an ever wider range of data collection methods. The development of telephone surveys in the 1990s offered a cheaper alternative to face-to-face interviews (Riandey and Firdion, 1993). Later, as the number of cell phone owners increased, technical advances made it possible to conduct detailed interviews and to directly select random samples of adults by means of cell phone numbers. More recently, the spread of the Internet has made it possible to conduct surveys with minimal data collection costs on "spontaneous" samples of volunteers or on representative samples of respondents recruited by phone. These innovations raise questions of sample representativeness and of the effect of data collection mode on the responses obtained. Analysing the Dutch version of an international survey on tobacco control, Mary THOMPSON *and her colleagues compare the results obtained on two subsamples, one that responded via the web, and the other by phone. They highlight differences linked to the respondents' characteristics and others attributable to the data collection mode, and present a method for estimating the results that would have been obtained using a single collection method, i.e. that used in the other countries.*

As obtaining a probability sample from a survey population becomes more and more difficult, survey practitioners are turning increasingly to mixed mode

* University of Waterloo, Canada.

** University of Amsterdam, the Netherlands.

*** University of Maastricht /CAPHRI, STIVORO for a smoke-free future, the Netherlands.

Correspondence: Y. Celia Huang, University of Waterloo, Department of Statistics and Actuarial Science, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1, tel.: + 1 519 888 4567 / 35543, email: ychuang@alumni.uwaterloo.ca

 DOI: 10.3917/pope.1303.0447

survey methods (Blyth, 2008). Telephone surveying is ideal for many purposes, because the questionnaire is administered by a person trained to elicit information and to keep the respondent engaged, and because travel costs are eliminated (Roberts, 2007). Random digit dial (RDD) methods (Groves et al., 1988) were developed to overcome the only serious issue with the frame in high penetration countries, namely the fact that many numbers are not listed. However, in recent years telephone frames have become less useful, partly because of the proportional increase in cell-phone-only households (Blumberg et al., 2006; Blumberg and Luke, 2008), and to an even greater extent because of access control technologies such as call display and automatic screening (Roberts, 2007; Tuckel and O'Neill, 2001).

One approach to mitigating the problem is to use additional frames and data collection modes in order to increase population coverage and response rates. Self-administered web data collection is particularly attractive because there are no interviewer or data entry costs (Blyth, 2008; Roberts, 2007). Thus, there is a great deal of interest in developing web survey frames and methods to encourage timely and good quality responses to surveys hosted on the web. Early on, web survey frames were mainly lists of email addresses gathered for other purposes. However, many survey firms (e.g. Harris Interactive and TNS NIPO) are developing databases and panels of people willing to respond to surveys on the web for appropriate compensation. The initial recruitment into the panel is often conducted by telephone or email, and the databases or panels can be described as "rich" frames because the respondent's personal data can be collected at recruitment. The combination of data collection modes presents new data-quality and analytic challenges (Frippiat and Marquis, 2010).

An immediate problem with combining telephone and web survey results, either across or within surveys, comes from the differences in the way questions are processed cognitively by respondents. For sensitive or difficult questions, or questions for which socially desirable responses exist, the presence or not of a human interviewer often makes a difference in how a respondent answers. For questions with a large number of response options, it is easier to choose with accuracy from a list which is seen than from a list which is heard. Thus, it is commonly observed that telephone respondents tend to give more socially desirable responses (Moskowitz, 2004; Kreuter et al., 2008), and more recently heard responses (Bishop et al., 1988). They are also more likely to choose the extreme ends of a 5-point Likert scale (Dillman et al., 2009). These kinds of phenomena can be called mode effects of the administrative sort or *administration effects*. Bowling (2005) has reviewed the effects on data quality of the mode of questionnaire administration.

Other types of mode effects found with mixed mode surveying can be called *selection effects*. These arise when the sample of respondents in one mode cannot be considered as a random subsample of the whole sample, as far as their characteristics are concerned. In some mixed mode surveys, respondents

are recruited from a single frame, and either assigned to a mode or allowed to choose between modes (De Leeuw, 2005). If respondents are allowed to choose, the subsamples for the two modes may ultimately differ on key characteristics. The other main kind of mixed mode design is a dual frame design, where respondents to the two modes are recruited from different but overlapping frames. Coverage by the frames, as well as non-response biases, may be expected to differ for the two modes (Nagelhout et al., 2010), again leading to different response distributions for the subsamples.

In simple comparisons of results from the two mode subsamples, administration and selection effects will be confounded. For example, if a higher proportion of a web subsample were to admit to a certain risky behaviour, we might suspect that this involves an administration effect in the sense that it is easier to admit such a thing when no interviewer is present; however, there might well be alternative explanations in terms of selection effects, associated with differences in the distributions of age or areas of residence in the subsamples.

In more complex comparisons, where variables associated with selection effects are controlled, it is more plausible to assume that remaining differences are administration effects. It may then be possible to model the administration effects when the telephone subsample and the web subsample are combined. This article illustrates an approach to such modelling that can be carried out with standard statistical software. The aim being to account for mode effects in the analysis of data from mixed mode surveys, it is not necessary to model the various sources of administration effects separately.

A useful concept for quantifying selection effects is the "propensity" to respond by one mode or the other (Rosenbaum and Rubin, 1983). Theoretically, this is the probability of responding by (say) telephone, given the fact of being in the combined sample of respondents, as a function of demographic variables *X,* and additional characteristics *W* which might influence the mode of response. In some applications, the propensity might be interpreted as the probability that the respondent chooses to respond by telephone, given a choice; in our application, it is the more simply the probability of having responded by telephone, conditional on having been contacted through one method or the other and having responded. It can be shown that, given a particular value of the propensity strictly between 0 and 1, the telephone and web parts of the sample are balanced with respect to the distribution of *X, W.* We cannot know the true propensity function, but we can approximate and estimate the propensity using a logistic regression model, regressing an indicator for responding by telephone on the covariates *X, W.* The resulting propensity score formula quantifies the selection effects which depend on these variables.(1)

(1) Note that if the compositions of the telephone and web samples were the same with respect to *X* and *W*, there would be no differential selection bias in those variables, and the propensity to respond by telephone would be estimated as a constant, namely the telephone sample size divided by the total sample size.

Controlling for propensity score in comparing the results from the two sample parts allows us in principle to separate administration mode effects from selection effects (i.e. from differential coverage and non-response bias). That is, if we compare telephone and web respondents with the same propensity score, the average mode differences will not be confounded with the variables X and W, and are therefore more likely to be administration effects.

When no assumptions are made concerning the directions of selection and administration effects, these effects are confounded, as pointed out by Vannieuwenhuyze et al. (2010) who advocate disentangling them by comparing a mixed mode data set with a corresponding single mode data set. In this article, we take the selection effect to be aligned with the propensity score, and the administration effect as the mode effect controlling for the propensity score. With these definitions, the two effects can both be estimated.

The responses may depend on covariates, possibly including those in X and W, other than through the propensity score. In that case, if we add covariates to the model, transformed to be orthogonal to the propensity score, we should still be able to interpret the coefficient of the propensity score as a selection effect. Where we omit such variables, as with the examples in Section III, then the estimated administration effect can be thought of as estimating the difference in response distribution – answering by telephone and web – for populations with the same joint distributions of X and W, as determined by conditioning on propensity score values.

In the final illustrative analyses presented in Section IV, we use the estimated propensity score not only to measure the selection effects, but also to control for them. If the set of variables in X and W seems likely to account for the selection effects and is not too extensive, simply controlling for X and W rather than for the propensity score will suffice.

When modelling responses that contain mode effects, the approach will surely depend on the purposes. However, in some cases it seems reasonable to take one mode to be the standard, with the effect of the other mode being characterized in terms of parameters in the model. For a specific outcome, the dependence on mode can then be expressed as (i) a dependence on X and W or the propensity score, quantifying the selection effects, and (ii) a transformation relating the patterns of responses in the standard mode to the patterns of responses in the other mode, quantifying the administration effects. In our illustration, using data from the International Tobacco Control Survey in the Netherlands (the ITC Netherlands Survey), the choice of the telephone mode as standard is arbitrary; it is not our intention to advocate a preference for one mode over the other, but to show how the results from both modes can be combined in a mixed mode design.

This article aims to suggest a way of modelling simultaneously the administration and the selection effects for outcomes measured on a five-point

ordinal scale, and a way of incorporating these effects into analyses in a multi-method study.

This approach to modelling can be used in several different ways: (i) to test for administration and selection effects in the response patterns for individual questions and to estimate their magnitude (as illustrated in this section); (ii) to test for common administration and selection effects in groups of questions; (iii) to use the administration effect parameters to "predict" the distribution of a respondent's response by telephone, given her/his response by web; and (iv) to account for collection mode effects when combining the web and telephone samples for analysis. As indicated, this article focuses on the first and last of these uses. In Section III, we show the results obtained when administration and selection effects are estimated by modelling for some questions in the ITC Netherlands Survey, and in Section IV, we give an example of a combined sample analysis.

This article is organized as follows. Section I introduces the data used for the model. In Section II the model is described in detail. In Section III the model is applied to selected questions from the ITC Netherlands Survey. Section IV presents the results of embedding the model in a cross-country comparison, and Section V is devoted to discussion.

I. The data

The International Tobacco Control Policy Evaluation Project (ITC) conducts longitudinal surveys, mainly of adult smokers, in 20 countries in order to evaluate policy measures being implemented under the World Health Organization Framework Convention on Tobacco Control (FCTC). In most of the countries, data collection is carried out either by telephone or face-to-face. However, mixed mode surveying has begun to enter the ITC Project. A description of the conceptual model of the ITC Project and the methods used in the earliest ITC surveys can be found in Fong et al. (2006) and Thompson et al. (2006).

The ITC Netherlands Survey, a survey of adult smokers with an oversampling of younger adults, differs from the other ITC surveys (prior to 2008) because most of the participants are responding to the CAWI (Computer-Assisted Web Interviewing) form of the questionnaire. TNS NIPO, the firm carrying out fieldwork in the Netherlands, has recruited an access panel (essentially a rich frame) of over 140,000 people from the general population for web surveys. The access panel is a non-probability panel recruited by TNS NIPO by phone or mail, but not by Internet. Since it is not possible to apply for participation, the panel has a relatively low number of "professional respondents", who participate in many web surveys as a way to generate income (Willems et al., 2006). Those invited to participate in the ITC Netherlands Survey constituted

a stratified random sample from the panel. Web surveying has become the preferred survey method in the Netherlands, as telephone surveying is not generally seen as cost-effective in that country and almost the entire population has Internet access (European Commission, 2008). It is a limitation of our analysis that the access panel is a non-probability panel, and we try to compensate for this fact in part by modelling the selection process.

In Wave 1, carried out between 13 March and 25 April 2008, the target was to recruit 1,700 CAWI participants aged 15 years or older, and more than 1,800 were obtained. The cooperation rate (the proportion of those invited and eligible who did respond, i.e. who answered the interview questions) was 78.0%. There was also an RDD (random digit dial) component of about 400 respondents aged 18 years or older, included for purposes of assessing the mode effects and facilitating comparison with the ITC surveys in France, Germany and the United Kingdom which are conducted entirely by telephone. For the RDD component, the cooperation rate was 78.1%. The response rate (the number of respondents as a proportion of the estimated number of telephone numbers attempted which could have reached eligible individuals) was only 4.2%. This is not unusual in the Netherlands, where response rates to telephone surveys have been declining since the 1990s (Bronner and Kuijlen, 2007). It should be noted that a low response rate does not necessarily translate into large non-response biases, particularly in econometric studies. Nagelhout et al. (2010) compared the demographic composition of the CATI (Computer Assisted Telephone Interviewing) sample with that of the population as determined by Statistics Netherlands (CBS), and found it to be reasonably representative.

The response distributions for many of the questions are different for CATI and CAWI administration, as shown by the Appendix Tables of this article and by some formal analyses in Nagelhout et al. (2008).

II. The model

The questions from the ITC Netherlands Survey chosen for analysis in Section III have five response options. We thus describe a model here for ordinal responses with five options. The basic model is the following, where Y denotes the coded response:

$$P(Y \leq d \,|u,\, \text{mode},\, X) = \frac{\exp(\eta_d)}{1+\exp(\eta_d)} \qquad [1]$$

where d goes from 1 to 4, and

$$\eta_d = c_d - (\alpha + \beta(d-1)) \times \text{mode} + b_0 u - B_1 Z \qquad [2]$$

Notice that the probability of the highest response, 5, is one minus the fourth probability given here.

To understand this ordinal logistic regression model [1], we can picture an imaginary underlying continuous response η for which the range is divided into five parts by response option thresholds $\eta_1 \leq \eta_2 \leq \eta_3 \leq \eta_4$. When η belongs to the dth part of the range, the observed response Y is equal to d. For example, if η is between η_3 and η_4, then Y is equal to 4, while if η is above η_4, then Y is equal to 5. The logit of the probability that $Y \leq d$ has a linear form in which the fixed explanatory variables are denoted by *mode* and Z, shown in [2]. The intercept parameters $c_1 \leq c_2 \leq c_3 \leq c_4$ can be thought of as location parameters for the response option thresholds. That is, a shift in c_d implies a corresponding shift in the expectation of the threshold η_d. In this model, the parameters α and β are the administration mode effect parameters. If the variable *mode* takes values 0 and 1, for web and telephone respectively, α represents the amount by which the telephone mode translates the locations, and β represents an amount by which the locations may be spread apart or contracted by the telephone mode. If there is a tendency for more extreme responses with telephone (mode = 1), then we would expect α to be negative (increasing η_1 and the probability for response 1) and β to be positive, with $3\beta > |\alpha|$ or $3\beta + \alpha > 0$ (decreasing η_4 and increasing the probability for response 5). If there is a tendency to select more recent responses with telephone, we would expect α to be non-significant or positive, and β again to be positive, leading to a decrease in all of the η_d. The (α, β) parameterization is intended as a parsimonious expression for the combined administration effects of mode.

The variable Z is the logit of the individual's estimated propensity to respond by telephone (mode 1) in terms of the covariates of interest X (such as sex, age group and education) and additional variables W; it is obtained from a separate logistic regression. Thus B_1, the coefficient of Z in the ordinal response model, is taken to be the selection effect parameter. Depending on the context, other variables explanatory of the response itself could be added, as in the example in Section IV. If they were not orthogonal to Z, the coefficient of Z would no longer be taken to measure the selection effect. However, within the model, the parameters α and β are still interpreted as administration effect parameters.

The variable u is a latent variable or random effect, which is assumed to be $N(0,1)$, independently for each individual, and b_0 is a positive multiplier. This variable is included to account for individual variability, and to allow the model to be fit using SAS PROC NLMIXED, where the presence of a random effect is required for convergence.

Useful references for ordinal response models like the one proposed here include McCullagh and Nelder (1989), and Grilli and Pratesi (2004).

III. Results for a selection of questions in the ITC Netherlands Survey

Before modelling the mode effects in the questions from the ITC Netherlands Survey, the first step in the method was to model the propensity to respond by telephone, using SAS PROC LOGISTIC. Web respondents under 18 years of age were removed from this illustrative analysis, since their telephone propensity would be 0. The variables *X* were taken to be sex, age group, and education, since these are demographic controls used in most ITC Project analyses. The additional characteristics *W* were marital status and some *individual attitude* variables – possible "webographic" variables in the terminology of Schonlau et al. (2007) – for which the response distributions had been found in a preliminary analysis to vary significantly by mode. These were:

- *time perspective* measured by the statement "You spend a lot of time thinking about how what you do today will affect your life in the future", with five response options "Strongly agree", "Agree", "Neither agree nor disagree", "Disagree", and "Strongly disagree", coded as 1 through 5;
- *personal stress #1* measured by "In the last six months, how often have you felt that difficulties were piling up so high that you could not overcome them?";
- *world event stress* measured by "In the last six months, how often have you been distressed by world events?";
- *personal stress #2* measured by "In the last 6 months, how often have you felt that you were unable to control the important things in your life?".

The latter three questions had five response options: "Never", "Almost never", "Sometimes", "Often" and "Very often", coded as 1 through 5. Individuals responding "Refused" and "Don't know" were coded as missing and excluded from the analysis. These individual attitude variables, treated as numeric, were included in the propensity model to capture possible sources of selection bias beyond those attributable to different distributions of sex, age and education in the two samples. For example, personality traits associated with preferred mode of communication may also be associated with one or more of the individual attitude variables.

When choosing the variables to include in the propensity formula, the aim is to produce a good predictor of response by telephone rather than to produce an explanation of it. If the propensity is well described by its model, controlling for the estimated propensity in the final term in the model can account for the sampling or selection effect of mode. Readers are referred to Rosenbaum and Rubin (1984) and Riou Franca et al. (2009) for further details of propensity score modelling.

The fitted propensity model, which was the basis for estimating individual propensity scores, is given in Table 1.

Table 1. Model for propensity for telephone response

Variables	Odds ratio (OR)	95% CI	*p*-value
Sex			
Female	0.90	(0.71 - 1.14)	0.3747
Male	1		
Age			< 0.0001[a]
18-24	0.53	(0.32 - 0.85)	0.0089
25-39	0.52	(0.37 - 0.72)	0.0001
40-54	0.94	(0.69 - 1.29)	0.6956
55+	1		
Educational level			0.0002[b]
High	1.67	(1.13 - 2.45)	0.0095
Medium	1.81	(1.37 - 2.39)	< 0.0001
Low	1		
Marital status			0.2157[b]
Married/common law/living together	0.78	(0.51 - 1.19)	0.2462
Single	0.65	(0.39 - 1.06)	0.0845
Divorced/widowed	1		
Lack of time perspective[c]	0.76	(0.67 - 0.86)	< 0.0001
Personal stress #1[c]	0.73	(0.63 - 0.85)	< 0.0001
World event stress[c]	1.40	(1.22 - 1.61)	< 0.0001
Personal stress #2[c]	0.85	(0.73 - 0.99)	0.0392

[a] Wald test with 3 degrees of freedom.
[b] Wald test with 2 degrees of freedom.
[c] See text for details.

Clearly the propensity to respond by telephone was much lower for those in the younger age groups, and although this finding partly reflects the fact that younger people are less likely to use landlines, it is largely due to the deliberate over-sampling of younger smokers from the web database. The propensity to respond by telephone was higher for those in the upper two education levels. Controlling for sex, age group, and education, the propensity to respond by telephone was lower for those scoring higher on "lack of time perspective" and the "personal stress" variables, and higher for those scoring higher on the "world event stress" variable. Translated into terms of selection effects, the results suggest that, relative to the telephone sample, the web sample over-sampled the younger age groups (as it did by design) and those with higher levels of lack of time perspective and of the personal stress variables. Also, the web sample under-sampled the upper two education levels and those with higher levels of the world event stress variable.[(2)]

As indicated earlier, the questions chosen for analysis from the ITC Netherlands Survey have five response options, such as "Never", "Rarely",

(2) If we had used web sample design weights which compensated for the over-sampling of younger age groups, the dependence of propensity on age group would not have been significant. However, it is actually useful for our illustrative purpose to have the propensity score associated with variables that might influence the responses.

"Sometimes", "Often" and "Very Often". The options are coded as 1 through 5. "Refused" and "Don't know" were recoded as missing.

The frequency tables of responses in the Appendix suggest that telephone respondents are more likely to select an extreme response; a more formal analysis can be found in Nagelhout et al. (2008).

Tables 2 to 4 show the estimates for the fitted ordinal response model for some individual questions. The questions belong to three groups of Likert items (questions with five ordinal response categories), grouped by subject matter and response options. The purpose of considering several questions per group is to examine whether the administration effects are consistent within a group. A summary of the results is that the α, β and $\alpha + 3\beta$ values are mainly consistent within groups, reflecting the patterns observed in the frequency tables. The Akaike Information Criterion (not shown), a commonly used measure of the relative goodness of fit, is improved in each case by the adding of α and β. The random effect scaling factor b_0 is not significantly different from 0 in any case. A propensity effect is present for many of the questions.

The first set, in Table 2, consists of questions that ask the respondent how often in the previous month, if at all, certain thoughts occurred.

Table 2. Model results for questions with five response options for frequency[(a)]

	How often in the previous month, if at all, did you think...							
	... about the harm your smoking might be doing to other people?		... about how much you enjoy smoking?		... about the harm your smoking might be doing to you?		... about the money you spend on smoking?	
Sample Sizes								
Web	1,578		1,576		1,579		1,581	
Phone	383		384		386		386	
	Estimate	(standard error)	Estimate	(standard error)	Estimate	(standard error)	Estimate	(standard error)
c_1	− 2.10	(0.43)	− 2.93	(0.14)	− 2.50	(0.13)	− 1.34	(0.11)
c_2	− 0.57	(0.16)	− 1.63	(0.12)	− 1.22	(0.11)	− 0.20	(0.11)
c_3	1.39	(0.28)	0.07	(0.11)	0.49	(0.11)	1.23	(0.11)
c_4	3.34	(0.50)	2.74	(0.15)	2.20	(0.14)	2.89	(0.14)
α	-0.92^{***}	(0.22)	-1.07^{***}	(0.16)	-0.55^{***}	(0.14)	-0.93^{***}	(0.12)
β	0.72^{***}	(0.14)	0.68^{***}	(0.07)	0.40^{***}	(0.07)	0.54^{***}	(0.06)
$\alpha + 3\beta$	2.32^{***}	(0.45)	2.00^{***}	(0.23)	0.66^{***}	(0.16)	0.69^{***}	(0.16)
b_0	0.71	(0.82)	0.00	(0.80)	0.01	(0.66)	0.01	(2.65)
B_1	0.54^{***}	(0.12)	0.32^{***}	(0.06)	0.46^{***}	(0.06)	0.03	(0.06)

[(a)] Response options: 1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, and 5 = Very often.
Note: $\alpha < 0$ and $\alpha + 3\beta > 0$ express greater tendency of telephone respondents than web respondents to give a low answer (1) or a high answer (5), respectively.
Significance levels: *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001.

The values of the administration mode effect parameters α, β and $\alpha + 3\beta$, are significant, similar across the set, and consistent with the possibility that telephone respondents give more extreme responses. The dependence on propensity to respond by telephone, as expressed by B_1, is statistically significant for the first three, and all estimates are positive, suggesting a closer to upper-end response for those more likely to respond by telephone (since the probabilities of the lower categories decrease as propensity rises); B_1 is not statistically significant in the last column. Other questions in the survey with the same set of response options have been analysed and shown to give similar though less strongly significant estimates for the α and β parameters; these results are not include here, to save space.

Table 3 shows results for questions in which respondents are asked to indicate the extent to which they agree with certain statements.

Table 3. Model results for questions with five response options for degree of agreement[(a)]

	Do you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree with the following statements?					
	Society disapproves of smoking		People who are important to you believe that you should not smoke		Smoking is sociable	
	Sample sizes					
Web	1,568		1,567		1,584	
Phone	383		384		386	
	Estimate	(standard error)	Estimate	(standard error)	Estimate	(standard error)
c_1	– 1.34	(0.12)	– 1.21	(0.11)	– 1.59	(0.53)
c_2	0.73	(0.11)	0.53	(0.11)	0.75	(0.31)
c_3	2.45	(0.13)	2.01	(0.12)	2.68	(0.81)
c_4	4.39	(0.19)	3.671	(0.16)	4.74	(1.17)
α	– 0.04	(0.14)	– 0.58***	(0.13)	– 0.59***	(0.22)
β	0.55***	(0.09)	0.31***	(0.08)	0.62***	(0.20)
$\alpha + 3\beta$	1.60***	(0.20)	0.35	(0.22)	1.28***	(0.27)
b_0	0.01	(1.65)	0.00	(0.67)	0.01	(.)
B_1	– 0.19**	(0.06)	– 0.18**	(0.06)	0.08	(0.06)

[(a)] Response options: 1 = Strongly agree, 2 = Agree, 3 = Neither agree nor disagree, 4 = Disagree, and 5 = Strongly disagree.
Significance levels: *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001.

The first variable in Table 3 is atypical, in that α is non-significant and barely negative, while β and $\alpha + 3\beta$ are significantly positive, reflecting a greater tendency to answer the most recent responses by telephone. The strong double negative in "disagree" and "disapproves" may be a factor in producing a different pattern. The second and third variables in Table 3 show results like those for Table 2, in line with a tendency for telephone respondents to select

more extreme responses, but with $3\beta + \alpha$ not significant for the second variable. For the first and second variables in Table 3 we see significant negative values for B_1. Since this increases the probabilities of the lower-end options, it appears that those with higher telephone propensity are less likely to choose the upper-end options (disagreement) with statements affirming disapproval of smoking, and perhaps are more sensitive to social norms. The results for β and B_1 suggest that, for the first question and to some extent the second, the sampling bias and the administration effects are pulling in opposite directions.

Table 4. Model results for questions with five response options [(a)]

	How easy would it be for you to quit smoking if you wanted to?		How much do you think you would benefit if you were to quit smoking permanently in the next 6 months?		If you decided to give up smoking completely in the next 6 months, how sure are you that you would succeed?	
Sample Sizes						
Web	1,572		1,546		1,554	
Phone	383		382		378	
	Estimate	(standard error)	Estimate	(standard error)	Estimate	(standard error)
c_1	– 2.86	(0.14)	– 3.17	(0.15)	– 0.72	(0.11)
c_2	– 0.99	(0.11)	– 0.98	(0.11)	0.78	(0.11)
c_3	0.22	(0.11)	0.48	(0.11)	1.97	(0.12)
c_4	1.26	(0.11)	1.60	(0.12)	2.74	(0.14)
α	– 0.22	(0.16)	– 0.52***	(0.16)	– 0.01	(0.12)
β	0.16*	(0.07)	0.18*	(0.08)	0.21**	(0.07)
$\alpha + 3\beta$	0.26*	(0.13)	0.01	(0.15)	0.61**	(0.19)
b_0	0.01	(.)	0.01	(.)	0.01	(2.89)
B_1	0.04	(0.06)	0.10	(0.06)	0.03	(0.06)

[(a)] Response options: 1 = Not at All, 2= Slightly, 3 = Moderately, 4 = Very much, and 5 = Extremely.
Significance levels: *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001.

The first variable in Table 4 shows a pattern for α and β similar to those of Table 2 and α is not significant. The results are again consistent with a tendency of telephone respondents to give more extreme responses. The second variable appears to show a tendency for telephone respondents to give closer to lower-end responses (due to an administration mode effect), and a non-significant tendency for those more likely (in terms of their characteristics) to respond by telephone to favour the response closest to the upper end. Interestingly, very similar proportions are obtained for the middle option[(3)] (33.95 for web and 34.90 for telephone); as well as a slightly greater use of the extreme options by telephone respondents. For the third variable in Table 4, the significantly positive β and $\alpha + 3\beta$ show telephone respondents more likely

(3) The corresponding frequency table can be viewed in Appendix 16 of the document http://www.itcproject.org/documents/researchmethods/appendixfrequencytablespdf

to choose the final, "extremely sure" option; the dependence on propensity is not significant.

Thus we see that the administration effects as we have expressed them are significant and fairly strong in all groups. They appear to be more similar within question groups where not only the set of response options but also the questions themselves are alike. For some questions a selection effect as well as an administration effect can be seen. It should be noted that we have expressed the administration effects only in terms of favouring responses close to one or both ends of the scale, and have not emphasized the possibility that they may include social desirability effects. Such effects may well be present, with size and direction depending on the nature of the question and the response options.

IV. Results for a label salience variable across four European countries

One interesting comparison across the ITC countries in Europe concerns the reactions of smokers to enhanced text warnings on cigarette pack labels introduced in the EU in 2003. The study of ITC results presented by Hitchman et al. (2011) suggests not only country-to-country differences in key variables, but also some difference in response patterns between the phone and web samples in the Netherlands. The other countries in the comparison are France, Germany and the UK, all with telephone as the mode of administration.

The *label noticing* variable: "In the last month, how often, if at all, have you noticed the warning labels on cigarette packages?" has five response options: "Never"; "Rarely", "Sometimes", "Often" and "Very often". Therefore, we applied the model of Section II, adding a country term, and including the propensity term only in the case of the Netherlands. The model is thus given by

$$P(Y \leq d \,|u, \text{mode}, X) = \frac{\exp(\eta_d)}{1+\exp(\eta_d)} \qquad [3]$$

where d goes from 1 to 4, and

$$\eta_d = c_d - (\alpha + \beta(d-1)) \times \text{mode} + b_0 u - B_1 Z\, I_{Neth} - B_2 f(X) - C\gamma \qquad [4]$$

For this model we again set the variable mode equal to 1 for telephone, and 0 for the web.

In formula (4), C is the set of country indicators, γ is the vector coefficient for the country indicators, I_{Neth} is an indicator for the Netherlands, and $f(X)$ represents a one-dimensional summary of the demographic variables which is a good predictor for noticing labels (reduced to accommodate a limitation of PROC NLMIXED). The demographic variables combined in the predictor

are sex, age group, ethnicity (country of birth for France), education, cigarettes per day and time to first cigarette. The variable Z stands for the logit of the propensity for web in the Netherlands. Cross-sectional survey weights, scaled to sum to country sample size, have been used in this analysis. The sample sizes are 1,383 for web and 347 for telephone in the Netherlands, and 1,559 in France, 1,361 in Germany and 1,412 in the UK. The results are given in Table 5. The reference level for country is Germany.

Table 5. Parameter estimates for a multi-country application of the mixed mode model to the question: "In the last month, how often, if at all, have you noticed the warning labels on cigarette packages?" [(a)]

Parameter	Estimate	Standard error	*p-value*
c_1	– 5.67	0.33	
c_2	0.21	0.28	
c_3	4.82	0.31	
c_4	10.98	0.39	
α	– 2.35	0.39	< 0.0001
β	2.05	0.13	< 0.0001
$\alpha + 3\beta$	3.81	0.45	< 0.0001
b_0	5.93	0.09	< 0.0001
B_1	0.54	0.20	0.0060
B_2	0.91	0.18	< 0.0001
France vs. Germany	3.30	0.28	< 0.0001
United Kingdom vs. Germany	3.46	0.26	< 0.0001
Netherlands vs. Germany	0.80	0.18	0.0493

[(a)] Response options: 1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, and 5 = Very often.

Table 5 shows α negative and significant, and β positive and significant, consistent with a tendency of telephone respondents to show more extreme responses. A significant and positive B_1 suggests that the higher the propensity to respond by telephone in the Netherlands, the higher the label salience, controlling for the covariates. The individual random effect coefficient b_0 is significant for this weighted analysis. The country effects follow the same pattern as seen in the analysis of Hitchman et al. (2011), where countries fell into two groups. Label salience was greatest, and about the same, for the United Kingdom and France. It was significantly lower in Germany and the Netherlands, but the difference between the Netherlands and Germany was barely significant at the 5% level.

The variable $f(X)$ in the Netherlands is not in fact quite orthogonal to Z in this example; the correlation between the two variables is 0.165 in the Netherlands. Therefore, the estimate of B_1 is not interpretable purely as a selection effect. When the model is fitted with no $f(X)$ term, the parameter

point estimates change very little, but their standard errors increase; the estimate of B_1 is reduced in magnitude to 0.4619 and its *p*-value becomes 0.0539; the *p*-value for Netherlands versus Germany becomes less significant at 0.1571. Thus the analysis with no $f(X)$ term is conservative, and the inclusion of $f(X)$ in the model means that the parameters are estimated with greater precision.

If we treat Netherlands telephone and Netherlands web as two separate countries, the model contains no terms for the administration or selection mode effects, but takes the form:

$$\eta_d = c_d + b_0 u - f(X)B_2 - C\gamma \qquad [5]$$

The results of model [5] are shown in Table 6.

Table 6. Parameter estimates for a multi-country application of the model without mode terms to the question: "In the last month, how often, if at all, have you noticed the warning labels on cigarette packages?" [(a)]

Parameter	Estimate	Standard error	*p-value*
c_1	– 3.47	0.45	
c_2	0.01	0.36	
c_3	3.30	0.41	
c_4	7.81	0.59	
b_0	6.19	0.42	< 0.0001
B_2	1.12	0.22	< 0.0001
France vs. Germany	3.69	0.35	< 0.0001
United Kingdom vs. Germany	3.88	0.36	< 0.0001
Netherlands telephone vs. Germany	– 0.086	0.42	0.8361
Netherlands web vs. Germany	– 1.06	0.27	< 0.0001

[(a)] Response options: 1= Never, 2 = Rarely, 3 = Sometimes, 4 = Often, and 5 = Very often.

Netherlands telephone is not significantly different from Germany, while Netherlands web is significantly different. The model of Table 5 expresses the same difference in terms of mode parameters, and allows the data from the two samples to be combined, while accounting for the mode effects.[(4)]

V. Discussion: summary and limitations

The results illustrate that the modelling approach presented in this paper can describe observed mode effects that appear to be administrative, at least for five-scale questions, and it can distinguish administration effects from

(4) Consistent with Table 6, if the two samples are combined, and the mode effects are *not* incorporated in the model, the Netherlands and Germany are separated in the label salience ranking, with coefficient estimate –0.87 and *p*-value 0.0004.

selection effects associated with collection mode. The model does not distinguish different types of selection effects, such as differential coverage bias and non-response bias, nor can it, without a richer collection of items, separate certain administration effects such as recency (the tendency to select the most recently heard response option) and social desirability bias if the most recently heard response option has greater social desirability. For five-scale questions, if the selection effects as defined in the models of this paper are not of interest, the administration effects can still be modelled as in the paper, with the control variables including those which would have been used in a propensity score model.

At the expense of making stronger assumptions, our method has an advantage over the one in Vannieuwenhuyze et al. (2010), in that no auxiliary single mode data set is needed.

To use the propensity score as a summary of variables on which differential sampling bias depends, it is necessary to have, at least conceptually, a large overlap in the coverage of the two frames. A different approach would have to be adopted in a situation where, for example, adults were surveyed by telephone and young people by web. For the ITC Netherlands Survey, the coverage of the telephone and web frames can be said to have a large overlap, although it is by no means complete: there are portions of the population without landlines or without Internet access, and hence without coverage by one frame or the other.

In the application to sets of questions in the ITC Netherlands Survey, we see that the administration effects are significant and fairly strong in all groups. The administration effects appear to be more similar within question groups where not only the set of response options but also the questions themselves are alike. The most consistent administration effect found in this study is that telephone respondents are more likely to select an extreme response than web respondents. This is consistent with previous research (Bronner and Kuijlen, 2007; Christian, Dillman, and Smyth 2005; Wichers and Zengerink, 2006). Respondents probably experience more time pressure on the phone and may use the extremes of a 5-point scale as if it were a yes/no scale. Primacy (a tendency to select the first-heard response option) and recency effects may also contribute to the extremity effect.

If other kinds of administration effects are of interest, other ways of parameterizing the model might be considered. However, to separate the sources of administration effects, the study would need to include questions and response options designed for this purpose; the ITC Netherlands questionnaire was not formulated with this aim in mind.

We might expect the estimated models to have different characteristics, because of cognitive processing differences, for questions which are four-scale (with no middle option), three-scale or binary. This could be a subject of future research. Note that if a question has a binary response, there is just one threshold point and it is not possible to identify both α and β; for binary

responses we could construct a simpler model, with terms for propensity and a mode indicator.

A limitation of our model is that it only includes respondents who selected one of the five response options, so it cannot describe an important mode effect, i.e. that web respondents tend to use the "don't know" option more than telephone respondents, (Bronner and Kuijlen, 2007; Roster, Rogers, and Albaum, 2004; Wichers and Zengerink, 2006) . This happens because web respondents see the "don't know" option listed on their computer screen, while telephone interviewers do not say that there is a "don't know" option. An earlier study showed that this was also the case for the ITC Netherlands Survey (Nagelhout et al., 2010).

Section IV provides one example, comparing response distributions across countries, of how the model presented here can be used in the analysis of data from the ITC surveys and other multi-country surveys. Similarly, we can compare response distributions within a dual frame design from one wave to the next, by dropping the country term and adding a term for wave. Note that the dependence from wave to wave of an individual's responses may be captured in the random effect term $b_0 u$. An analysis comparing changes in distributions over time among several countries can be carried out by adding terms for country, wave, and wave by country.

In summary, we believe the modelling approach presented in this article provides a natural and useful framework for accounting for mode of interviewing in mixed mode surveys. This is a relatively new topic. Mode effects are often tested for, but are only beginning to be incorporated in models.

Acknowledgements: The authors would like to thank the reviewers for their valuable comments and suggestions that helped improve the quality of the manuscript.

Appendix

Frequency tables for selected questions (CAWI and CATI modes)

Table A.1. "You spend a lot of time thinking about how what you do today will affect your life in the future"

Time perspective aDI42211	Mode = web (CAWI)				
	Frequency	% (excl. "don't know")	Cumulative frequency	Cumulative % (excl. "don't know")	Cumulative %
1. Strongly agree	113	6.50	113	6.50	6.21
2. Agree	448	25.78	561	32.28	30.82
3. Neither agree nor disagree	751	43.21	1,312	75.49	72.09
4. Disagree	323	18.58	1,635	94.07	89.84
5. Strongly disagree	103	5.93	1,738	100.00	95.49
Total	1,738	100.00	1,738	100.00	95.49
9. Don't know	82	–	1,820	–	100.00

Time perspective aDI42211	Mode = telephone (CATI)			
	Frequency	%	Cumulative frequency	Cumulative %
1. Strongly agree	58	14.36	58	14.36
2. Agree	165	40.84	223	55.20
3. Neither agree nor disagree	78	19.31	301	74.50
4. Disagree	78	19.31	379	93.81
5. Strongly disagree	24	5.94	403	99.75
9. Don't know	1	0.25	404	100.00

Table A.2. "In the last six months, how often have you felt difficulties were piling up so high that you could not overcome them?"

Personal stress #1 aDI42326	Mode = web (CAWI)				
	Frequency	% (excl. "don't know")	Cumulative frequency	Cumulative % (excl. "don't know")	Cumulative %
1. Never	366	20.77	366	20.77	20.11
2. Almost never	670	38.02	1,036	58.79	56.92
3. Sometimes	530	30.08	1,566	88.87	86.04
4. Often	151	8.57	1,717	97.44	94.34
5. Very often	45	2.55	1,762	100.00	96.81
Total	1,762	100.00	1,762	100.00	96.81
9. Don't know	58	-	1,820	-	100.00

Stress personnel #1 aDI42326	Mode = telephone (CATI)			
	Frequency	%	Cumulative frequency	Cumulative %
1. Never	183	45.30	183	45.30
2. Almost never	93	23.02	276	68.32
3. Sometimes	88	21.78	364	90.10
4. Often	29	7.18	393	97.28
5. Very often	11	2.72	404	100.00

Table A.3. "In the last six months, how often have you been distressed by world events?"

World event stress aDI42331	Mode = web (CAWI)				
	Frequency	% (excl. "don't know")	Cumulative frequency	Cumulative % (excl. "don't know")	Cumulative %
1. Never	202	11.44	202	11.44	11.10
2. Almost never	574	32.50	776	43.94	42.64
3. Sometimes	806	45.64	1,582	89.58	86.92
4. Often	159	9.00	1,741	98.58	95.66
5. Very often	25	1.42	1,766	100.00	97.03
Total	1,766	100.00	1,766	100.00	97.03
9. Don't know	54	–	1,820	–	100.00

World event stress aDI42331	Mode = telephone (CATI)			
	Frequency	%	Cumulative frequency	Cumulative %
1. Never	65	16.09	65	16.09
2. Almost never	69	17.08	134	33.17
3. Sometimes	185	45.79	319	78.96
4. Often	66	16.34	385	95.30
5. Very often	18	4.46	403	99.75
8. Refused	1	0.25	404	100.00

Table A.4. "In the last six months, how often have you felt that you were unable to control the important things in your life?"

Personal stress #2 aDI42331	Mode = web (CAWI)				
	Frequency	% (excl. "don't know")	Cumulative frequency	Cumulative % (excl. "don't know")	Cumulative %
1. Never	258	14.72	258	14.72	14.18
2. Almost never	660	37.65	918	52.37	50.44
3. Sometimes	602	34.34	1,520	86.71	83.52
4. Often	189	10.78	1,709	97.49	93.90
5. Very often	44	2.51	1,753	100.00	96.32
Total	1,753	100.00	1,753	100.00	96.32
9. Don't know	67	–	1,820	–	100.00

Personal stress #2 aDI42331	Mode = telephone (CATI)			
	Frequency	%	Cumulative frequency	Cumulative %
1. Never	118	29.21	118	29.21
2. Almost never	115	28.47	233	57.67
3. Sometimes	129	31.93	362	89.60
4. Often	28	6.93	390	96.53
5. Very often	13	3.22	403	99.75
9. Don't know	1	0.25	404	100.00

See website: http://www.itcproject.org

REFERENCES

BISHOP George F., HIPPLER Hans-Juergen, SCHWARZ Norbert, STRACK Fritz. 1988, "A comparison of response effects in self-administered and telephone surveys", in GROVES Robert M., BIEMER Paul P., LYBERG Lars E., MASSEY James T., NICHOLLS II William L., WAKSBERG Joseph (eds.), *Telephone Survey Methodology,* New York, John Wiley, pp. 321-340.

BLUMBERG Stephen J., LUKE Julian V., 2008, "Wireless substitution: Early release of estimates based on data from the National Health Interview Survey", www.cdc.gov/nchs/nhis.htm

BLUMBERG Stephen J., LUKE Julian V., CYNAMON Marcie L., 2006, "Telephone coverage and health survey estimates: Evaluating the need for concern about wireless substitution", *American Journal of Public Health,* 96(5), pp. 926-931.

BLYTH Bill, 2008, "Mixed modes: The only 'fitness' regime?", *International Journal of Market Research*, 50(2), pp. 241-266.

BOWLING Ann, 2005, "Mode of questionnaire administration can have serious effects on data quality", *Journal of Public Health,* 27(3), pp. 281-291.

BRONNER Fred, KUIJLEN Ton, 2007, "The live or digital interviewer: A comparison between CASI, CAPI and CATI with respect to differences in response behavior", *International Journal of Market Research,* 49(2), pp. 167-190.

CHRISTIAN Leah Melani, DILLMAN Don A., SMYTH Jolene D., 2005, "The effects of mode and format on answers to scalar questions in telephone and web surveys", *Second International Conference on Telephone Survey Methodology*, Miami, USA.

DE LEEUW Edith D., 2005, "To mix or not to mix data collection modes in surveys", *Journal of Official Statistics,* 21(2), pp. 233-255.

DILLMAN Don, PHELPS Glenn, TORTORA Robert, SWIFT Karen, KOHRELL Julie et al., 2009, "Response rate and measurement differences in mixed-mode surveys: Using mail, telephone, interactive voice response and the internet", *Social Science Research,* 38(1), pp. 1-18.

EUROPEAN COMMISSION, 2008, "E-communications household survey", European Commission, Brussels.

FONG Geoffrey T., CUMMINGS K. Michael, BORLAND Ron, HASTINGS Gerard, HYLAND Andrew et al., 2006, "The conceptual framework of the International Tobacco Control (ITC) Policy Evaluation Project", *Tobacco Control*, 15 (Suppl. III), pp. iii3-iii11.

FRIPPIAT Didier, MARQUIS Nicolas, 2010, Web surveys in the social sciences: An overview", *Population, English Edition*, 65(2), pp. 285-312.

GRILLI Leonardo, PRATESI Monica, 2004, "Weighted estimation in multilevel ordinal and binary models in the presence of informative sampling designs", *Survey Methodology,* 30(1), pp. 93-103.

GROVES Robert M., BIEMER Paul P., LYBERG Lars E., MASSEY James T., NICHOLLS II William L., WAKSBERG Joseph (eds.), 1988, *Telephone Survey Methodology,* John Wiley & Sons, 608 p.

HITCHMAN Sara, MONS Ute, NAGELHOUT Gera E., GUIGNARD Romain, MCNEILL Ann et al., 2011, "Effectiveness of the European Union text-only cigarette health warnings: findings from four countries", *European Journal of Public Health*, 22(5), pp. 693-699.

KREUTER Frauke, PRESSER Stanley, TOURANGEAU Roger, 2008, "Social desirability bias in CATI, IVR, and Web surveys: The effects of mode and question sensitivity", *Public Opinion Quarterly*, 72(5), pp. 847-865.

MCCULLAGH Peter, NELDER John A., 1989, *Generalized Linear Models*, Second Edition, Chapman and Hall, 532 p.

MOSKOWITZ Joel M., 2004, "Assessment of cigarette smoking and smoking susceptibility among youth: Telephone computer-assisted self-interviews versus computer-assisted telephone interviews", *Public Opinion Quarterly,* 68(4), pp. 565-587.

NAGELHOUT Gera E., WILLEMSEN Marc C., THOMPSON Mary E., FONG Geoffrey T., VAN DEN PUTTE Bas, DE VRIES Hein, 2008, "Mixed modes in the International Tobacco Control (ITC) Netherlands Survey: Differences between CAWI and CATI on smoking-related questions", Joint conference of SRNT and SRNT-Europe, April 27-30, 2009, Dublin, Ireland.

NAGELHOUT Gera E., WILLEMSEN Marc C., THOMPSON Mary E., FONG Geoffrey T., VAN DEN PUTTE Bas, DE VRIES Hein, 2010, "Is web interviewing a good alternative to telephone interviewing? Findings from the International Tobacco Control (ITC) Netherlands Survey", *BMC Public Health*, 10(1), pp. 351-360.

RIANDEY Benoît, FIRDION Jean-Marie, 1993, "Vie personnelle et enquête téléphonique, L'exemple de l'enquête ACSF", *Population*, 48(5), pp. 1257-1280.

RIOU Franca L., DAUTZENBERG Bertrand, FALISSARD Bruno, REYNAUD Michel, 2009, "Are social norms associated with smoking in French university students? A survey report on smoking correlates", *Substance Abuse Treatment, Prevention, and Policy*, 4(4), 14 p.

ROBERTS Caroline, 2007, "Mixing modes of data collection in surveys: A methodological review", *NCRM Methods Review Papers, Volume NCRM/008* London, Centre for Comparative Social Surveys, City University.

ROSENBAUM Paul R., RUBIN Donald B., 1983, "The central role of the propensity score in observational studies for causal effects", *Biometrika,* 70(1), pp. 41-55.

ROSENBAUM Paul R., RUBIN Donald B., 1984, "Reducing bias in observational studies using subclassification on the propensity score", *Journal of the American Statistical Association,* 79(387), pp. 516-524.

ROSTER Catherine A., ROGERS Robert D., ALBAUM Gerald, 2004, "A comparison of response characteristics from web and telephone surveys", *International Journal of Market Research*, 46(3), pp. 359-373.

SCHONLAU Matthias, VAN SOEST Arthur, KAPTEYN Arie, 2007, "Are 'Webographic' or attitudinal questions useful for adjusting estimates from Web surveys using propensity scoring?", *Survey Research Methods*, 1(3), pp. 155-163.

THOMPSON Mary E., FONG Geoffrey T., HAMMOND David, BOUDREAU Christian, DRIEZEN Pete et al., 2006, "Methods of the International Tobacco Control (ITC) Four Country Survey", *Tobacco Control*, 15 (Suppl III), pp. iii12-iii18.

TUCKEL Peter, O'NEILL Harry, 2001, "The vanishing respondent in telephone surveys", Proceedings of the Annual Meeting of the American Statistical Association.

VANNIEUWENHUYZE Jorre, LOOSVELDT Geert, MOLENBERGHS Geert, 2010, "A method for evaluating mode effects in mixed-mode surveys", *Public Opinion Quarterly*, 74(5), pp. 1027-1045.

WICHERS Bart, ZENGERINK Evelien, 2006, "It's the culture, stupid! A cross-cultural comparison of data collection methods", Amsterdam, ESOMAR.

WILLEMS Pieter, BROWN Millward, VAN OSSENBRUGGEN Robert, VONK Ted, 2006, "The effects of panel recruitment and management on research results", Dutch online panel study NOPVO, Barcelona, ESOMAR.

Mary E. Thompson, Y. Celia Huang, Christian Boudreau, Geoffrey T. Fong, Bas Van Den Putte, Gera E. Nagelhout, Marc C. Willemsen • **Accounting for the Effects of Data Collection Method. Application to the International Tobacco Control Netherlands Survey**

Mixed mode surveys are becoming increasingly common. This has led to calls for tests of the differences in response patterns between survey modes. In this article, we present an analysis of mode effects, using data from Wave 1 of the ITC Netherlands Survey, conducted by web (CAWI) and telephone (CATI). For many of the questions, the web and telephone samples differed in the distribution of response options. This was found to be partly attributable to selection effects, since the web and telephone respondents were recruited in different ways, and the web and telephone samples differed on demographic characteristics. Another source of difference in the response option distribution was "administrative" in origin, having to do with the tendency of respondents to process the options differently depending on survey mode. This article illustrates an approach to modelling in a mixed mode survey that takes into account both selection and administration mode effects. The model is also embedded in an analysis of reactions to labels on cigarette packages using ITC data from the Netherlands, Germany, France, and the United Kingdom.

Mary E. Thompson, Y. Celia Huang, Christian Boudreau, Geoffrey T. Fong, Bas Van Den Putte, Gera E. Nagelhout, Marc C. Willemsen • **Mesurer les effets de la méthode de collecte des données. Application à l'enquête ITC des Pays-Bas**

Les enquêtes recourant à un mode mixte de collecte sont de plus en plus nombreuses, et il devient nécessaire de tester les écarts entre les réponses obtenues par téléphone et par internet. Cet article analyse les effets des différents modes de collecte en utilisant la vague 1 de l'enquête International Tobacco Control (ITC) des Pays-Bas réalisée par internet (CAWI) et par téléphone (CATI). Pour de nombreuses questions, les échantillons présentent des distributions de réponse différentes. C'est dû en partie à des effets de sélection, car les répondants sont recrutés par des procédures différentes et les échantillons n'ont pas les mêmes caractéristiques démographiques, et en partie à des facteurs d'administration des questions, les répondants traitant différemment les items de réponse en fonction du mode de collecte. L'objectif est ici de présenter une modélisation qui prend en compte à la fois les effets de sélection et d'administration dans une enquête utilisant un mode mixte de collecte. Le modèle est aussi intégré dans une analyse des réactions à des mentions figurant sur les paquets de cigarettes, d'après les enquêtes ITC aux Pays-Bas, en Allemagne, en France et au Royaume-Uni.

Mary E. Thompson, Y. Celia Huang, Christian Boudreau, Geoffrey T. Fong, Bas Van Den Putte, Gera E. Nagelhout, Marc C. Willemsen • **Medir los efectos del método de colecta de datos. Aplicación a la encuesta holandesa Internacional Tobacco Control**

Los modos mixtos de colecta son cada vez más frecuentes en las encuestas, y se hace necesario comprobar las diferencias entre las respuestas obtenidas por diferentes métodos. Este artículo analiza los efectos del modo de colecta utilizando los datos de la primera ola de la encuesta holandesa International Tobacco Control (ITC), realizada por internet (CAWI) y por teléfono (CATI). Para muchas preguntas, las dos muestras presentan distribuciones de repuestas diferentes. Ello es debido en parte a efectos de selección pues los sujetos que respondieron por teléfono fueron escogidos de manera diferente a los que respondieron por internet, y también al hecho de que las dos muestras diferían por ciertas características demográficas. Además intervino otro factor de origen "administrativo", pues hubo una tendencia a tratar las opciones de respuesta diferentemente según el modo de colecta. Este artículo presenta una modelización que toma en cuenta ambos efectos, selección y administración, en una encuesta con un modo mixto de colecta. El modelo es utilizado también en un análisis de las reacciones a las etiquetas que figuran en los paquetes de cigarrillos en Holanda, Alemania, Francia y Reino-Unido.

Keywords: mixed mode survey, mode effects, propensity score, ordinal responses.