Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2005 : Methodological Challenges for Future Information needs

2005

Statistics Canada
Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2005
Methodological Challenges for Future Information Needs

INCORPORATING TIME-IN-SAMPLE IN LONGITUDINAL SURVEY MODELS

M. E. Thompson, C. Boudreau, and P. Driezen[1]

ABSTRACT

Time-in-sample effects include rotation group biases, memory errors changing over time, and panel conditioning. It is clearly important to take such possibilities into account when designing and analyzing longitudinal survey data before and after an intervention. The question arises of how these effects can be separated from effects of attrition and the true impacts of intervention. It is suggested that using comparable control groups and newly recruited sample at each wave should help with modeling and identification of time-in-sample effects. We illustrate this with data from the International Tobacco Control Four Country Survey, a longitudinal survey of smokers in four countries in which tobacco control interventions are being introduced.

KEY WORDS: Longitudinal surveys; Time-in-Sample; Cohort design

1. INTRODUCTION

1.1 Definition and Examples

At the time of recruitment, the participants in a longitudinal survey are chosen to be representative of population. As time goes on, typically some of the participants will drop out, and dropout may be informative in the sense of depending on the response variables of interest. However, even if dropout is minimal, the participants who continue to the second and third waves of a longitudinal survey may differ from those they supposedly represent in subtle ways, leading to what are known as time-in-sample effects.

In a longitudinal or panel survey or cohort study, a time-in-sample effect could be defined as an effect of a respondent's prior reporting on that respondent's current or future responses; it is a dependence of response characteristics on the length of time or the number of times a respondent has been observed.

This phenomenon has long been known. Bailar (1975) noted a rotation group bias in the United States Current Population Survey, in which participants were under observation for four months, out of the sample for eight months, and under observation again for four months. Unemployment estimates by cohort or rotation group tended to be higher for groups at the beginning of either of their four-month rotation periods. Ghangurde (1982) reported a similar phenomenon for rotation groups in the Canadian Labour Force Survey, in which participation is for 6 consecutive months. Again, unemployment estimates by rotation group were highest for the group most recently recruited.

Perhaps the first explanation that comes to mind might be that unemployment rates are higher among those more likely to drop out of the survey as the months go on. However, an examination of the data in both the United States and Canada suggested that there was also some other influence at work. Differences between rotation groups tended to persist after adjustments for non-response and attrition.

[1] M. E. Thompson, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1 (methomps@uwaterloo.ca); C. Boudreau, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1 (cboudrea@uwaterloo.ca); P. Driezen, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1 (prdrieze@uwaterloo.ca)

It is well known that in nutrition surveys the mean reported intake of various dietary components tends to decrease with time in sample. Thus an adjustment is used to reconcile measures at different waves. Furthermore, the variances of intake measures tend to decline. Nusser et al (1996) used a transformation approach to estimating usual daily intake. Their initial regression adjustment modified the data to remove sequence effects in the mean of the intake distributions for successive survey days. Because of the possibility of higher-order time-in-sample effects, they were led to standardize the sample variance for survey days after the first to the sample variance observed on the first day.

A paper by Wilson and Howell (2005) examined arthritis trends in the longitudinal Health and Retirement Study (HRS) in the United States:

> "In this study, we show that (1) age-specific arthritis prevalence in the longitudinal Health and Retirement Study (HRS) … has risen sharply since its inception in 1992, and (2) this rise is almost surely spurious. In periods for which the data sets are comparable, we find no such increase in the cross-sectional National Health Interview Survey (NHIS) … even though prevalence in the HRS rises sharply between 1992 and 1996 for 55-56 year olds, the prevalence for that age group plummets to its 1992 level among the new cohort added in 1998 and then rises rapidly again between 1998 and 2002."

The authors were able to demonstrate that the discrepancies were not due to sample attrition in the HRS by determining a lower bound on the upward trend, assuming participants lost to follow-up maintained their last seen disease state, i.e. did not contribute to the upward trend. The authors' explanation was that

> "Participation in an extensive health survey sensitizes the participants to be more concerned for their health, and consequently, to seek more additional medical attention than they would otherwise."

Thompson (2004) noted the possibility of similar phenomena in connection with the National Population Health Survey (NPHS) of Statistics Canada. The NPHS longitudinal sample incidence of seeking contact with alternative medicine providers increased rapidly over the first three cycles. Furthermore, 1998 self-reported relapse rates of women who were smokers in 1994 and who had stopped smoking by 1996 were unexpectedly low, of the order of about 50%. Each of these trends could be explained in part by increased awareness resulting from participation in the survey.

1.2 Origins and Impacts of Time-in-Sample Effects

Various explanations of time-in-sample phenomena have been brought forward, depending on the context. Respondents may try to make their responses at the current wave consistent with previous responses, in what has been called the "Socratic effect" (Jagodzinski et al, 1987). Respondents may become fatigued, and disengage from thinking about responses. Respondents may learn from prior interviews that certain responses are less socially desirable, or lead to increased probing, and may change their subsequent responses strategically (Bailar, 1975). Even apart from the impact of the influences being studied, participation in a panel survey may cause changes in attitudes and behaviour, referred to as "panel conditioning" (Duncan and Kalton, 1987). For some kinds of questions, recall may improve from one cycle to the next; perhaps "in the last 30 days" or "in the last 24 hours" is easier to construe the second or third time one is surveyed. An interesting analysis is provided by Biderman and Cantor (1984).

The impacts on analyses include most fundamentally time-in-sample bias in means, and decreases in variability of responses. But time-in-sample effects may also interfere with the determination of the effects of aging, as in the arthritis study, or the effects of changes in the environment, as with the ITC study discussed in the next section. Differentiating time-in-sample effects from the effects of sample attrition is a delicate matter.

2. ITC FOUR COUNTRY SURVEY

2.1 Survey Design

The flagship survey of the International Tobacco Control Policy Evaluation Project is the ITC Four Country Survey, led by Professor Geoffrey Fong of Psychology at the University of Waterloo (Fong et al, 2006). This is an international longitudinal survey of approximately 2000 smokers in each of Canada, the United States, the United Kingdom, and Australia. The waves are approximately annual: Wave 1 was carried out in late 2002, and Wave 4 in late 2005. The main objective of the survey is to evaluate the impact and effectiveness of new tobacco control measures by comparing behavioral changes in the country where a new policy is being introduced to behavioral changes (or lack thereof) observed by other countries where no such measure is in place. For example, we have been able to look at the effects of an advertising ban in the United Kingdom by comparing changes in the United Kingdom sample of smokers to the ones in the other three countries before and after the ban (Harris et al, 2006).

The sampling and data collection have used RDD frames, and the sampling design for smoker households approximates stratified random sampling with proportional allocation. Retention per wave has varied between 80% and 60%, being consistently lowest in the U.S. Respondents lost to follow-up have been replenished with a fresh cohort at each wave, with the same sampling scheme as used at the beginning. The replenished sample is weighted to prevalence figures from national benchmark surveys, and the weighted and pooled sample of smokers appears to be reasonably representative not only in terms of age and sex, but also in terms of cigarette consumption.

The cohort design is imperfect: the sample sizes for cohorts after the first are determined by attrition rather than by a precision analysis. However, one of the reasons for adopting this design came from experience with another similar survey, still in progress, in which time-in-sample effects for some key variables were strongly suspected.

2.2 Responses over Time

We have examined the course of key variables in the ITC data over the first three waves. First, plots like those in Figures 1-3 were used to get a general idea of potential time-in-sample effects.

For Figure 1, the question put to respondents in each wave, pertaining to the previous six months, was the following:

> Now I would like you to think about advertising or information that talks about the dangers of smoking, or encourages quitting. In the last 6 months – since [6M anchor] – how often, if at all, have you noticed such advertising or information? *(read out response options)*
> 01 – Never
> 02 – Rarely
> 03 – Sometimes
> 04 – Often
> 05 – Very often

There were no national anti-smoking campaigns in the observation period, though there were undoubtedly campaigns in sub-national jurisdictions, which will eventually be systematically documented.

The "cross –sectional samples", shown in the solid lines, show proportions "often or very often" which decrease and then increase. Cohort 1 consists of respondents present in all three waves. Except in Australia, Cohort 1 shows a steeper decline between Waves 1 and 2, but stays approximately constant from Wave 2 to Wave 3. Cohort 2 consists of respondents recruited at Wave 2 who are also present at Wave 3. Except in the U.S., Cohort 2 starts at approximately the same level as Cohort 1 and declines in the same manner. Thus it appears that noticing declines with time in sample. Could it be that this can be explained by higher dropout rates by those who have noticed messages more? The solid line position at Wave 2 is a weighted average of the Cohort 2 level, the Cohort 1 level, and the level (not shown) for those Wave 2 participants not included in either cohort. Since the solid line position and the Cohort 1 position are close, it would appear that those dropping out after Wave 2 are those who are noticing less often at Wave 2. Similarly, the solid line and the Cohort 2 position are close at Wave 3.

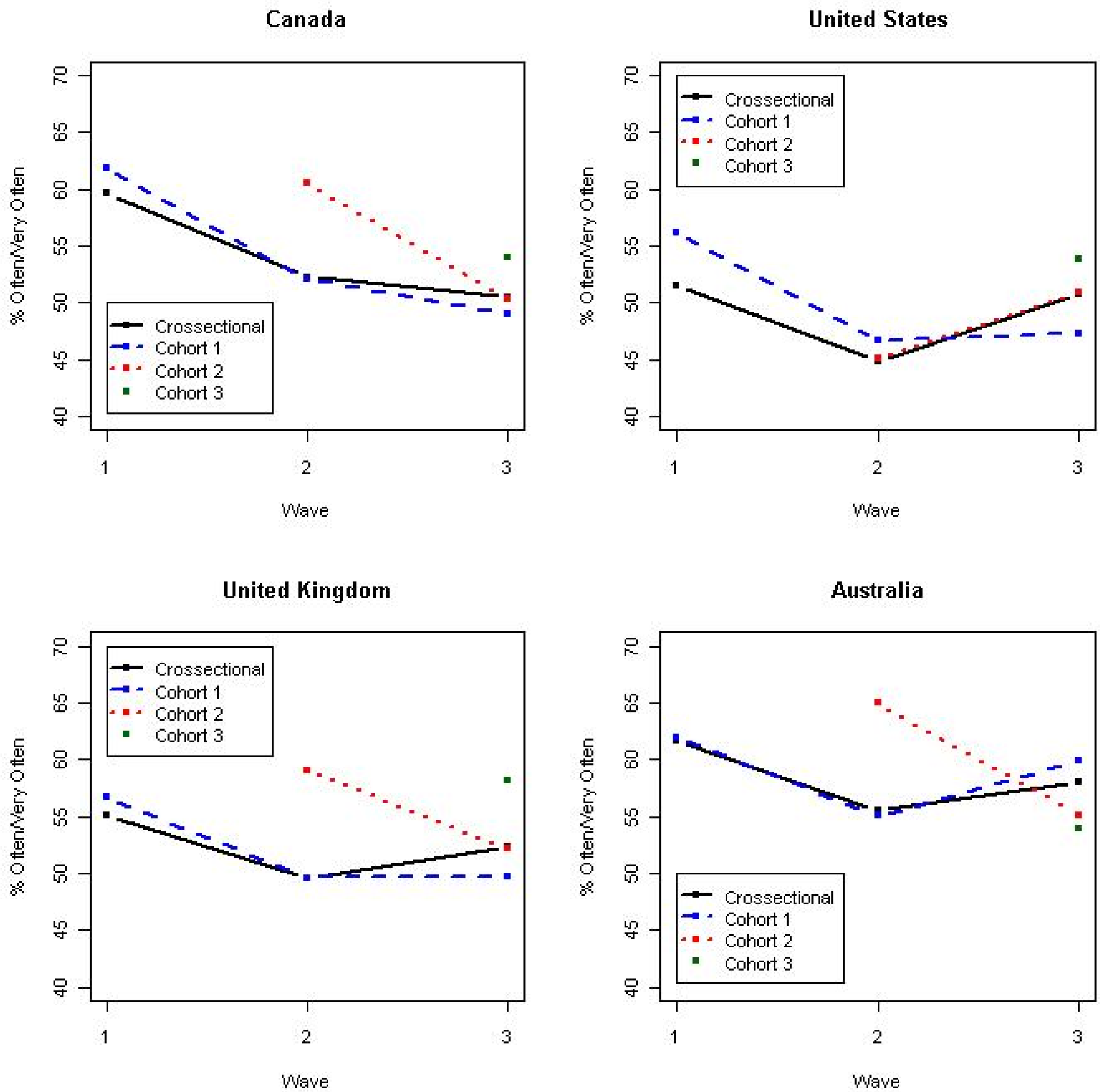

Figure 1: Percentages noticing information about dangers of smoking, by country and cohort

For this question a general estimating equations (GEE) analysis shows a significant effect for a "time-in-sample" covariate.

This kind of pattern is not seen for all variables. For example, there is no downward or upward trend for self-perceived health. However, it is fairly common for "noticing" variables in the ITC survey. Figure 2 shows that in all four countries there was a decline in the noticing of promotion of smoking. The decline between Waves 1 and 2 was (significantly) steepest in the UK, where there actually was an advertising ban. And again, we see a tendency for the newer cohorts to be noticing more, in the absence of intervention.

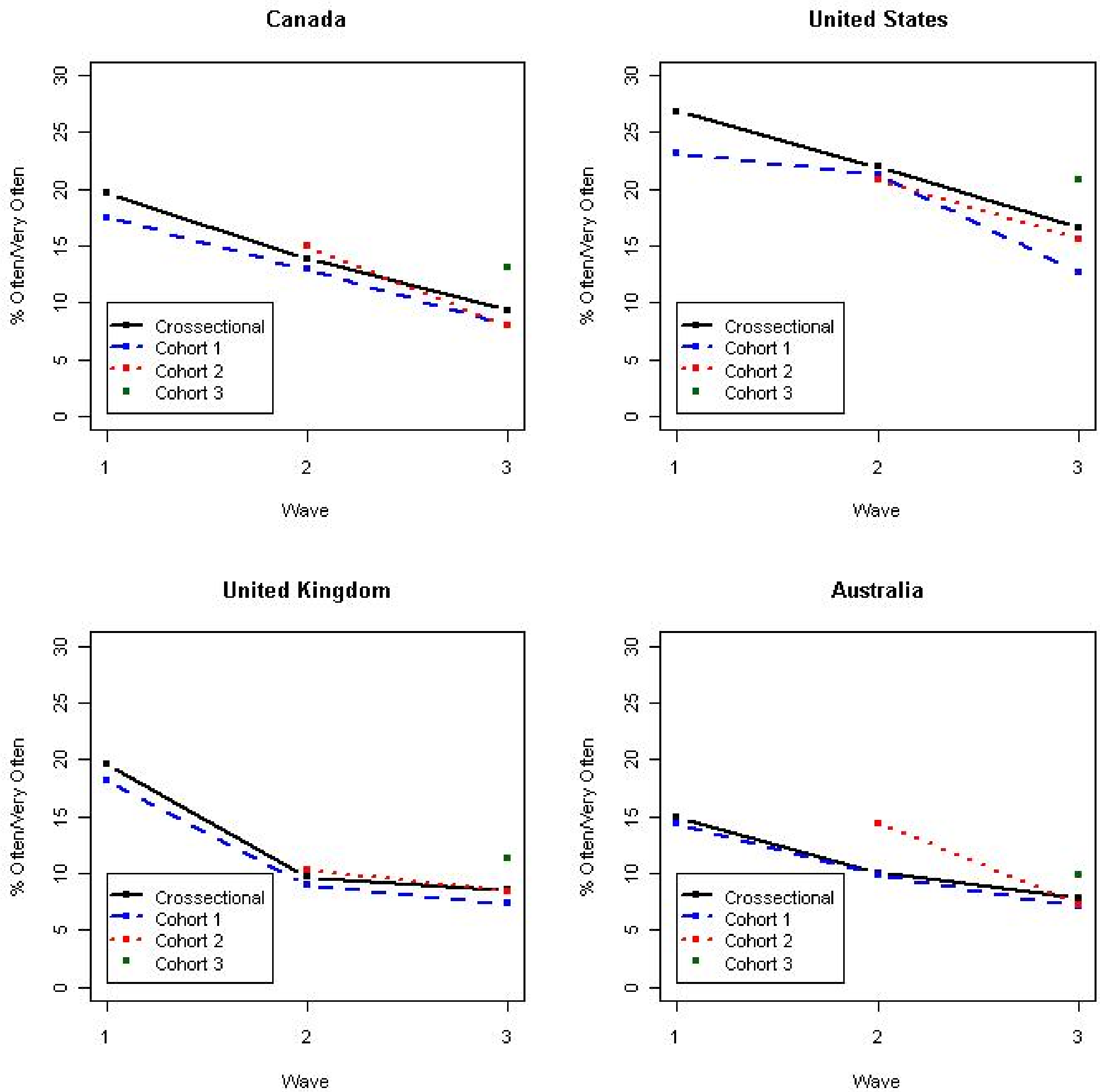

Figure 2: Percentages noticing promotion of smoking, by country and cohort

Figure 3 shows that the noticing of warning labels tended to decline within a cohort, except in the UK where new warning labels were introduced, between Waves 1 and 2.

The general findings from our analysis are that some variables appear to show a time-in-sample effect, while others do not; to some extent the design helps us to separate time-in-sample from attrition. The countries appear qualitatively consistent for the most part; "noticing" appears to go down with time in sample.

Is this effect important? For the ITC Project the effect is certainly quite important. The investigators are trying to evaluate the impact of policy changes addressing a serious and long-standing public health problem, and one way is to try to understand how various kinds of messages influence knowledge, beliefs, and the intention to quit smoking. Noticing is both an outcome and a potential mediating variable between policy and changes in beliefs and behaviour. We need to be able to measure noticing, or a useful surrogate for noticing.

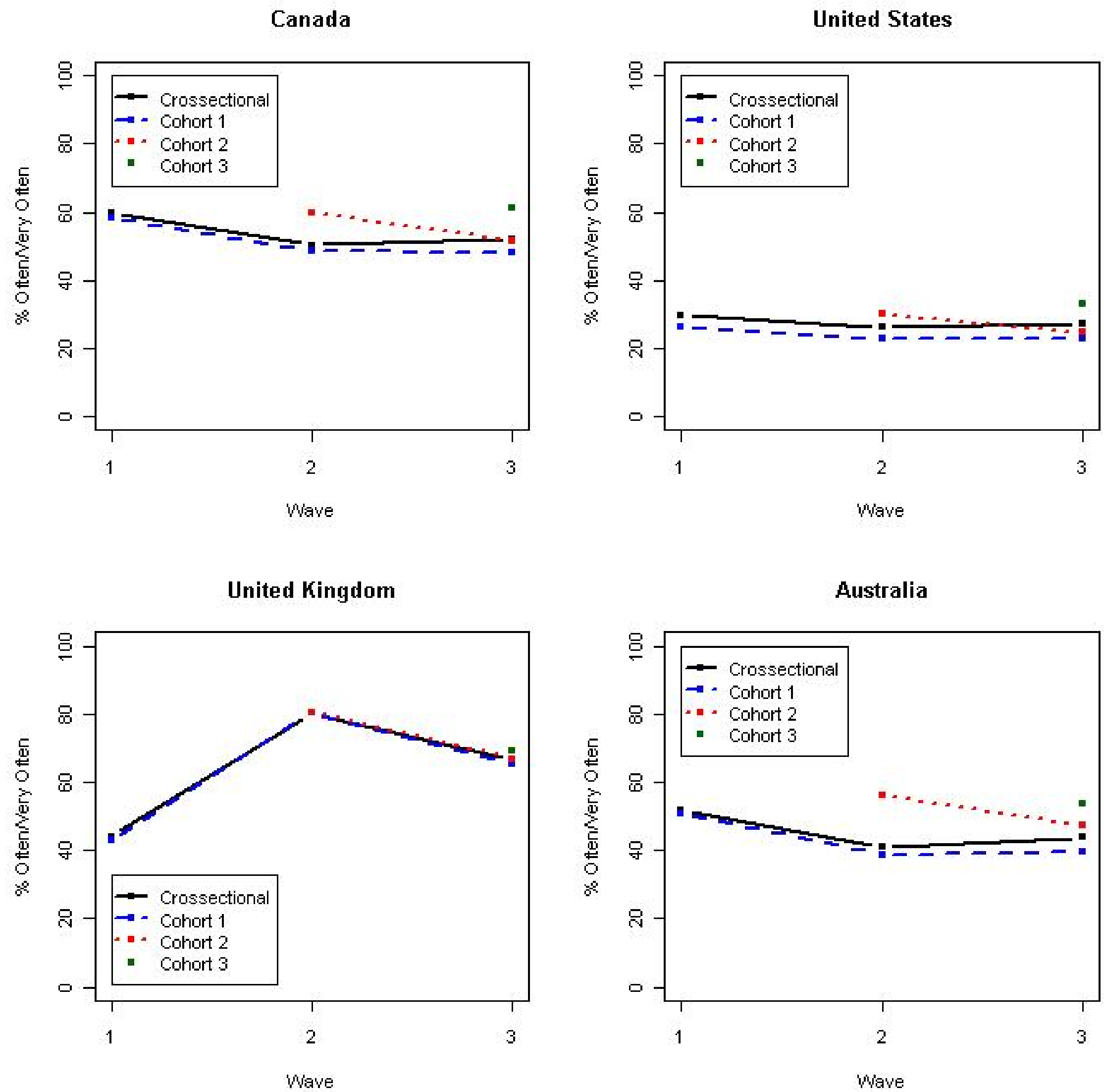

Figure 3: Percentages noticing warning labels in the last month, by country and cohort

We looked at possible explanations for the effects that we seem to be seeing. They are counterintuitive, since we might expect noticing to increase once the respondent has been questioned about the messages. The answer does not appear to lie in strategies for completing the interview quickly, as will be seen in the next section. The noticing questions are often gateways to other modules, but these are not long or tedious. The most likely explanation seems to be a different cognitive processing of the recall question the second and third times around.

2.3 Models Incorporating Time-in-Sample Effects

In longitudinal data we can incorporate a time-in-sample covariate into a GEE model in a straightforward manner, even using survey weights with standard software. However, GEE models are marginal, and they measure the association of an individual's successive responses through covariance modeling. Moreover, they only approximate internally consistent probability models, and they do not handle well attrition that is not completely random. In the study of time-in-sample effects, it is also of interest to consider growth curve models with random coefficients. For continuous outcomes we would use the general form of a linear mixed effects model for a vector outcome $Y_i = (Y_{i1},\ldots,Y_{ip})'$, representing repeated observations for an individual *i:*

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \text{ where } b_i \sim N(0,\sigma_b^2) \quad \text{and } \varepsilon_i \sim N_p(0,\sigma^2 I_p).$$

With the ITC data we used R to fit a model to the logarithm of the time taken to complete the interview, using a linear mixed effects model. Besides time-in-sample, the fixed effect covariates were country, sex, age and ethnicity; potentially the individual intercept and slope for time in sample were random, although we eventually chose only the intercept to be random. The results are shown in Table 1. The coefficient of time-in-sample is significant but small, showing a slight increase in interview length with time in sample.

Table 1: Results of fitting a growth curve model to the logarithm of interview completion time

	$\hat{\beta}$	se $(\hat{\beta})$	p-value		$\hat{\sigma}$	95% CI
(Intercept)	3.5687	0.00905	< 0.0001	(Intercept)	0.0929	(0.0864, 0.0999)
Time-in-sample	0.000045	0.00001	< 0.0001	Within	0.2402	(0.2368, 0.2437)
Country:			< 0.0001†			
Australia	0					
Canada	0.1025	0.00661	< 0.0001			
United Kingdom	0.0166	0.00651	0.0110			
United States	0.0845	0.00796	< 0.0001			
Age	0.0027	0.00018	< 0.0001			
Sex:						
Female	0					
Male	0.0128	0.00492	0.0093			
Ethnicity	0.0581	0.01591	0.0003			

† 3 d.f. test

We could have included indicators for attrition patterns to refine the analysis further. That is, we could have included an indicator for dropout after a single wave, an indicator for dropout after two waves, and indicators for the three cohorts of the plots in the previous section.

By using numerical methods for binary and ordinal random effects models, we could address the noticing questions in a similar manner. The covariates would include indicators for the interventions.

3. SUMMARY

In longitudinal surveys, time-in-sample effects may be common; the associated measurement biases may be hard to disentangle from aging effects, intervention effects, and attrition effects. We can try to design an observational study to tease these apart, by observing parallel populations with and without interventions, and by introducing fresh cohorts at each wave. Models which incorporate time-in-sample and attrition pattern indicators should be useful in isolating intervention effects.

ACKNOWLEDGMENTS

This research was supported in part by NSERC. The ITC Four Country Survey is funded by grants from the U.S. National Cancer Institute/NIH (from the Roswell Park Transdisciplinary Tobacco Use Research Center (TTURC), P50 CA111236, and from R01 CA100362), the Canadian Institutes for Health Research (57897), Robert Wood Johnson Foundation (045734), the Australian National Health and Medical Research Council (265903), Cancer Research UK (C312/A3726), the Australian Commonwealth Department of Health and Ageing, the Centre for Behavioural Research and Program Evaluation of the National Cancer Institute of Canada/Canadian Cancer Society, and the Canadian Tobacco Control Research Initiative.

REFERENCES

Bailar, B. (1975), "The Effects of Rotation Group Bias on Estimates from Panel Surveys", *Journal of the American Statistical Association,* 70, pp. 23-30.

Biderman, A. D. and Cantor, D. (1984), "A Longitudinal Analysis of Bounding, Respondent Conditioning, and Mobility as Sources of Panel Bias in the National Crime Survey. *Proceedings of the Section on Survey Research Methods of the American Statistical Association.*

Duncan, G. J., and G. Kalton (1987), "Issues of Design and Analysis of Surveys Across Time", *International Statistical Review,* 55, pp. 97-117.

Fong, G. T., K. M. Cummings, R. Borland, G. Hastings, A. Hyland, G. A. Giovino, D. Hammond, and M. E. Thompson (2006), "The Conceptual Framework of the International Tobacco Control Policy Evaluation Survey", submitted paper.

Ghangurde, P. D. (1982), "Rotation Group Bias in the LFS Estimates", *Survey Methodology,* 8, pp. 86-101.

Harris, F., A. M. Mackintosh, S. Anderson, G. Hastings, R. Borland, G. T. Fong, D. Hammond, K. M. Cummings (2006), " The Effects of the Advertising/Promotion Ban in the United Kingdom on Awareness of Tobacco Marketing: Findings from the ITC Four Country Survey", *Tobacco Control.*

Jagodzinski, W., S. Kuehnel and P. Schmidt (1987), "Is there a `Socratic Effect' in Non-Experimental Panel Studies? Consistency of an Attitude Toward Guestworkers", *Sociological Methods and Research,* 15, pp. 259-302.

Nusser, S. M., A. L. Carriquiry, K. W. Dodd, and W. A. Fuller (1996), "A Semi-parametric Transformation Approach to Estimating Usual Daily Intake Distributions", *Journal of the American Statistical Association,* 91, pp. 1440-1449.

Thompson, M. E. (2004), Uses of the National Population Health Survey in examining Health Behaviour Changes, with Particular Reference to Smoking Cessation.
http://tdr.tug-libraries.on.ca/SWORDCSITE/DOCS/Rsummary04.html

Wilson, S. E., and B. L. Howell (2005), Do Panel Surveys Make People Sick? Arthritis Trends in the Health and Retirement Study", *Social Science and Medicine,* 60, pp. 2623-2627.