



THE ONTARIO  
TOBACCO  
RESEARCH  
UNIT

UNITÉ  
DE RECHERCHE  
SUR LE TABAC  
DE L'ONTARIO

# **The Future of Household Telephone Surveys on Tobacco: Methodological and Contextual Issues**

Thomas Stephens  
K Stephen Brown  
David Ip  
Roberta Ferrence

June 2009

## **Acknowledgement**

The authors would like to acknowledge the contributions made by presenters at the concurrent session, *The Future of Household Telephone Surveys on Tobacco: Methodological and Contextual Issues*, whose expertise has been invaluable:

Susan Bondy, University of Toronto  
Christian Boudreau, University of Waterloo  
Lori Diemert, Ontario Tobacco Research Unit  
Murray Kaiserman, Health Canada  
Diane Kunyk, University of Alberta  
Julie McAuley, Statistics Canada  
David Northrup, York University  
Julie Snider, Health Canada  
Mary E. Thompson, University of Waterloo  
Charles Victor, Ontario Tobacco Research Unit.

Suggested Citation: Stephens T, Brown KS, Ip D, Ferrence R. *The Future of Household Telephone Surveys on Tobacco: Methodological and Contextual Issues*. Toronto, ON: The Ontario Tobacco Research Unit, Special Report Series. June 2009.

## Table of Contents

Acknowledgement.....	ii
Table of Contents .....	iii
Executive Summary.....	1
Background.....	1
Introduction.....	4
Background to the Working Session .....	4
Objectives of the Working Session .....	4
Objectives of this Report.....	5
Declining Response Rates: An Old Challenge Revived .....	6
Response Rates and their Components.....	6
Some Reasons for the Decline in Response Rates.....	7
Coverage/Response Issues and Bias.....	8
Are Low Response Rates Always a Problem? .....	9
Some Responses to these Challenges.....	9
Innovative Approaches to Improving Response Rates .....	10
Innovative Approaches to Dealing with Non-response.....	11
Measurement Issues: A Secondary, but Enduring Challenge.....	13
Cigarette Sales as a Measure of Consumption .....	13
Question Order and Context.....	13
Measuring Youth Smoking.....	14
Use of Standard Definitions .....	14
Next Steps: New Initiatives to Consider.....	15
Collaborative Initiatives .....	15
Developmental Initiatives .....	15
Educational/Publishing Initiatives.....	15
Conclusions.....	16
Current Practices.....	16
Future Directions.....	16
Appendix A: Speaker Summaries of Invited Presentations .....	18
What’s Past is Prologue Even with Tobacco Use Surveys.....	19
Definitions and Changing Patterns of Response Rates.....	24
Measurement Issues in Tobacco Control .....	27
A Tobacco-Control Advocate Reflects on the Use of Survey Data.....	30
Representativeness of Federal Cross-Sectional (Tobacco) Surveys.....	33
Longitudinal Surveys and Attrition: Issues and Solutions.....	36
Methodological Issues in Reporting Cigarette Use and Consumption.....	40

Comparisons of School-Based and Telephone Surveys of Youth .....	45
Accounting for the Effects of Data Collection Method in Surveys Collecting Data Using Both the Web and the Telephone.....	51
Declining Response Rates: The Experience of the International Tobacco Control Policy Project .....	55
Appendix B: Participant Questions for the Final Discussion Session.....	60

## **Executive Summary**

### **Background**

Household telephone surveys provide an important tobacco-control tool for research, monitoring and evaluation, and policy-making, but declining response rates of surveys and various measurement issues are undermining the quality and utility of survey data. In November 2008, the Ontario Tobacco Research Unit (OTRU) hosted the Ontario Tobacco Control Conference: *Tobacco Control for the 21<sup>st</sup> Century: Challenges in Research and Evaluation*. Approximately 30 survey researchers and data users met in a concurrent working session to consider the state of telephone surveys of tobacco use. Presentations and discussion considered the decline in response rates and selected measurement issues. The participants made many suggestions to address these issues, and identified some promising developments. This report highlights the issues discussed and recommendations made at the concurrent working session.

### **The Issues**

Issues affecting the generalizability of tobacco-control research now and in the future include:

- With declines in smoking prevalence, it is increasingly difficult for household telephone surveys to locate and capture smokers.
- Response rates have decreased, especially within large urban areas; obtaining a representative sample of smokers by telephone requires a larger sample and is more costly.
- Growing household smoking bans may make smokers with only land-line phones less accessible.
- Non-standardized definitions and measures often affect the validity and reliability of results of survey research.

### **Recommendations and Future Directions**

A series of recommendations and future directions were identified:

#### **Survey Design and Planning**

- Conduct fewer, but smarter, surveys (more targeted, less repetition).
- Establish a coordinated ethics review process that is common across institutions and is consistent across the country.
- Adapt Random Digit Dialing (RDD) sampling frames to include cell and internet phones.
- Develop sampling frames for conducting surveys on the web.

- Create pre-screened panels of potential survey participants.
- Develop, test and refine multi-mode surveys, especially web-based surveys to complement telephone surveys.
- Consider (again) what variables should be monitored beyond tobacco use.
- Use standard definitions for key constructs across surveys.
- Develop and implement a standard definition of “adults” in surveys.
- Develop better coordination between federal and provincial players with respect to designing and conducting general population and school surveys.
- Develop a national data base of school-level smoking rates.
- Establish an ongoing working group on response rate issues, with a focus on developing a respondent panel.

### **Field Operations**

- Develop methods to encourage timely and accurate responses to surveys on the web.
- Place more onus on call-bank providers to remove ineligible numbers and to include cell and cable-based phones.
- Use biochemical validation to assess standard questions on smoking behaviour.
- Develop new methods for reaching populations that currently are under-represented.

### **Research and Analysis**

- Identify the parameters that affect representativeness within existing survey data sets (e.g., in-school status) in order to adjust the sample through re-weighting.
- Conduct sensitivity analyses of the impact on survey estimates of various approaches to re-weighting and imputation.
- Look for opportunities to combine data from separate, compatible surveys.
- Conduct a thorough empirical examination of the factors behind declining response rates as a guide to developing effective solutions.
- Examine the association between survey context/question order and estimates of key indicators including smoking status.

### **Knowledge Transfer**

- Document best practices for maintaining/increasing response rates and share this with survey practitioners and funding agencies.
- Prepare a commentary on how to interpret and describe response rates and share this with researchers, advocates, policy-makers, editors and funding agencies.
- Report data for youth (less than age 20) by single year of age where possible.

### **Other Solutions**

- Use available regulatory and legal mechanisms to improve the quality of data on cigarette sales, by clearly distinguishing retail from wholesale sales, and sales containing contraband and smuggled products.

### **Conclusions**

There was a general consensus that the *status quo* threatens to undermine the quality and usefulness of survey data. Household telephone surveys, with their declining response rates, are becoming less representative. While conventional strategies may not solve the problem, identified recommendations provide possible next steps to work towards solutions.

## Introduction

### Background to the Working Session

In November 2008, the Ontario Tobacco Research Unit (OTRU) hosted the Ontario Tobacco Control Conference: *Tobacco Control for the 21<sup>st</sup> Century: Challenges in Research and Evaluation*. The meeting was designed to provide a forum for setting priorities and developing recommendations for research and evaluation in tobacco control.

The conference was built around a series of keynote speakers and nine concurrent working sessions.<sup>1</sup> The working session that is the focus of this Special Report is *The Future of Telephone Surveys on Tobacco: Methodological and Contextual Issues*. At the session, which ran for 10 hours over two days, invited speakers made brief remarks (see Appendix A), and there was considerable discussion among the 25-30 persons present. Participation was by prior registration, which ensured both interest in the topic and some continuity of participation.

### Objectives of the Working Session

Household survey data are an important tool for tobacco control, used for research, monitoring and evaluation, and policy-making, and they have a long, if somewhat checkered, history in Canada.<sup>2</sup> However, methodological and contextual issues increasingly affect the representativeness of surveys and the portrayal of smokers in household surveys of the general population. Many of these issues are particularly salient for telephone surveys, which have become the standard means of reaching households. These issues potentially limit our understanding of smokers' perceptions and behaviours.

With declines in smoking prevalence, it is becoming more difficult for household telephone surveys to locate and capture smokers. Response rates have also decreased, especially within large urban areas, so that obtaining a representative sample of smokers by telephone requires a larger sample and more resources than it used to. With more households banning smoking indoors, it may also be harder to reach smokers using only land-line telephones. These issues have implications for the generalizability of tobacco control research for the next decade and beyond.

<sup>1</sup> The conference program can be found online: <http://www.otru.org/conference/pdfs/PAAG.pdf>.

<sup>2</sup> Ferrence R, Stephens T. Monitoring Tobacco Use in Canada: The Need for a Surveillance Strategy (Commentary). *Chronic Diseases in Canada* 2000; 21: 50-53.

The focus of the working session was methodological and contextual issues related to the quality of data from tobacco-use surveys, particularly surveys employing the telephone. The objective was to share information on the issues, get a better understanding of the scope and nature of the problem, identify actual and potential solutions, and lay the groundwork for some constructive future developments.

Roberta Ferrence outlined the key issues with tobacco-related surveys, and Tom Stephens described the long history of tobacco surveys in Canada. Using Ontario examples, David Northrup illustrated the definitions of response rates and described their changing patterns. Sue Bondy used two different perspectives to consider the response rate issue, while Diane Kunyk reflected on using survey data for advocacy. The first afternoon session mostly focused on the *implications of non-response for survey representativeness*. Judy Snider and Julie McAuley discussed cross-sectional federal tobacco surveys. Charles Victor and Lori Diemert examined issues and solutions relating to longitudinal surveys and attrition. Murray Kaiserman explained some of the methodological issues in the reporting of cigarette use and consumption.

*Dealing with Non-Response* was the theme for the morning session of Day 2. K. Stephen Brown compared school-based and telephone-based methods for collecting tobacco-use data from youth; Kelli-An Lawrance described using web surveys on university campuses. Mary Thompson examined the effects of collecting survey data on the web and by telephone. Christian Boudreau shared his experience with declining response rates from the International Tobacco Control Policy surveys. The afternoon session was devoted to an open discussion where participants and presenters considered key questions and issues related to the previous three sessions (see Appendix B), discussed possible answers and solutions, and developed strategies for next steps.

## **Objectives of this Report**

This report brings together the main points raised in the working session, with a focus on the many discussions. It includes material from the presentations, but does not cover all the points raised by the presenters. For this detail, see Appendix A for the presentation summaries.

Although there was much discussion in the working session and even some recommendations, this report should not be interpreted as describing a consensus among participants. However, this OTRU Special Report does serve as a record of the event and will hopefully also stimulate further developments in this important area.

## Declining Response Rates: An Old Challenge Revived

### Response Rates and their Components

Response rates have always been a key consideration in population surveys. At a minimum, good rates lower the cost per person contacted, contribute to obtained sample size, and, perhaps unfairly, raise perceptions of the quality of the survey data, and may enhance publication prospects. Poor response rates achieve the opposite. Perhaps even more importantly, response rates can affect the representativeness of the survey sample if they are not uniform among groups of interest.<sup>3</sup>

Compared to in-person interviews of the general population, telephone surveys face additional challenges. This is because they typically employ random-digit dialing (RDD) to reach eligible respondents, and non-contact becomes an issue along with refusal to participate.

Refusals are the main challenge to response rates in cross-sectional surveys. For example, in the current Centre for Addiction and Mental Health (CAMH) Monitor, which is a typical telephone survey conducted by an experienced university-based research institute, refusals account for 80% of total non-response, and only 20% is due to a failure to contact (Northrup, Appendix A). In contrast, non-contact is the main component in later waves of longitudinal surveys. For example, in the Ontario Tobacco Survey, only 2% of respondents lost to follow-up were due to a refusal at Wave 2 (Victor & Diemert, Appendix A); in the International Tobacco Control (ITC) surveys, refusal was only 2-7% in later waves while non-contact grew from 27% to 50% over the course of three waves (Boudreau, Appendix A).

Overall, current response rates for cross-sectional surveys by reputable field organizations are typically around 50%. Some are much lower, depending on the population, field organization, sponsor, and other factors. For example, rates in the ITC surveys have been as low as 5% in the Netherlands for the telephone component (Thompson, Appendix A).

There has been a general steady trend downward in response rates for some years; it is experienced in both the private and public sectors, and across nations and survey topics. Prior to 2000, the decline was about 2-3% per year; some figures suggest that the annual decline has more than doubled recently (Boudreau, Appendix A).

<sup>3</sup>If the “groups of interest” are defined by Census-type characteristics such as age, gender, education level, or region of residence, post-data-collection adjustments to the survey weights can restore the correct proportions in the sample and remove any related bias. However, if the groups of interest are defined by less visible traits such as smoking status, attitudes toward smoke-free spaces, propensity to quit, and so on, response bias will not only be difficult or impossible to adjust, it may not even be detected.

## **Some Reasons for the Decline in Response Rates**

While the Working Session focused on the growing non-response issue and devoted a considerable amount of time to current and potential ways to address the problem, there was comparatively little time or discussion of the reasons for the decline in response rates. Many observations, suggestions, and hypotheses were advanced, but there was relatively little data-based explanation of the decline in rates. A thorough examination of the issue, via empirical studies, is recommended as a logical and necessary step in developing effective solutions.

Among the reasons suggested for the decline in response rates are the following:

### ***Telephone Access Issues***

Since the telephone is the sole means of accessing sample respondents in the surveys under consideration here, the nature and distribution of telephone access to households is critical. At present, the banks of telephone numbers supplied to survey firms are composed of exchanges based exclusively on conventional land-line services. Cell phones and cable-based exchanges (VOIP) are not part of current phone banks, while a large number of non-residential and fax numbers *are* included. By December 2007, over six percent of Canadian households were served by cell phones *only*, a substantial increase since 2003 (McCauley & Snider, Appendix A). Another, unspecified, proportion are served only by cable, not phone, companies; these households are also excluded from current phone banks. Finally, even if cell phones and VOIP service were included in RDD phone banks, call display and call blocking are available for telephones of all kinds, and these are presumed to be used increasingly to avoid survey calls. However, there is no evidence (yet) that call-management technology is having an adverse impact on survey response.

### ***Ethics Board Constraints***

The requirement for ethical approval faced by institutional-based survey research poses some additional challenges. Ethics boards may limit the number of calls for converting those who initially refuse to take the survey to participants, and may prohibit calling cell-phone exchanges, because of the cost to the called party. They can impose privacy-related constraints on follow-up in longitudinal surveys and demand active rather than passive consent. School-based surveys, which may also have a telephone component, can involve multiple levels of consent and additional ethics review, which can consume time and provide more opportunities for refusal (Brown, Appendix A).

### ***Economic and Social Climate***

While the incidence of telemarketing and opinion polls has not been quantified, an apparent increase has led to public calls for do-not-call registries. This increase appears to have produced a certain sense of saturation on the part of the public, even when the cause is worthy. A sense of being used on the part of would-be survey participants and a general decline of deference and civility in society combine to discourage survey response. Added to this may be survey-specific issues, such as fatigue

or boredom with detailed smoking questions, especially in longitudinal surveys such as the 1994 Survey of Smoking in Canada (SoSiC) series (Stephens, Appendix A).

### **Coverage/Response Issues and Bias**

While non-response is always a concern, it does not *necessarily* undermine sample representativeness and lead to bias. As noted above, a low response rate will increase survey costs and may reduce the size and power of the sample, but it will not otherwise affect the quality of survey data -- unless the rate of non-response is unevenly distributed in the sample. Unfortunately, “MNAR” (missing not at random) is the most frequent reason for attrition in follow-up studies: dropouts are disproportionately related to the “outcome” of interest, such as smoking status in the OTS (Victor & Diemert, Appendix A).

More generally, if non-participation is related to sample characteristics under study, bias will result. Such an effect would only exacerbate the long-standing under-coverage of certain hard-to-reach populations in surveys—newer immigrants, non-English- or French-speakers, the homeless, students, the mentally ill, and the institutionalized.

Several examples of potential bias due to low response rates were discussed:

- Some groups are more likely than others to have only a cellular phone (and thus not to be reached by RDD): people with low income, those living in large urban centers or Western Canada, renters and people living with roommates, men in the trades, transient populations, students, and people age 15-24, particularly males (McCauley & Snider, Appendix A). While under-coverage of some of these groups can be compensated for with adjustments to the data collection strategy or survey weights, this can only work with Census-type individual characteristics.
- In school/youth surveys, parents required to give active consent are known to be different from passive-consent parents. This may lead to bias if consent procedures are not uniform throughout such a survey.

While web-based surveys may be used to supplement telephone surveys (see below, *Innovative Approaches to Improving Response Rates* (page 10)), not all internet users are equal. For example, smokers who are more likely to use the internet tend to be younger, richer, more educated, in better health, and have less psychological distress.<sup>4</sup>

<sup>4</sup>Stoddard JL, Augustson EM.. Smokers who use internet and smokers who don't: data from the Health Information and National Trends Survey (HINTS). *Nicotine & Tobacco Research* 2006 Dec; 8 Suppl 1:S77-85.

## **Are Low Response Rates Always a Problem?**

It is clear that low response rates can lead to increased survey costs, possibly reduced sample size, and potential bias. Higher costs and smaller samples obviously are problems; whether or not bias is an issue is not so clear-cut.

Whether non-response bias is a genuine problem depends largely on the intended use of the survey data. Descriptive studies, such as monitoring and population-based evaluation, require representative, reliable and recent data. Advocacy is also dependent on data free of obvious bias and other low-response issues (Kunyk, Appendix A). In contrast to this perspective, modelers are less concerned about representativeness. Indeed, for research on mechanisms, the least representative sample may be the *most efficient* (Bondy, Appendix A). However, journal reviewers and editors are not likely to be as concerned with cost-efficient data collection as with large and representative samples resulting from high response rates (notwithstanding the fact that journals are full of studies on small samples of carefully selected clinical patients).

Regardless of the intended use of the data, survey response rates may be approaching a threshold below which they will be dismissed as necessarily inconclusive and unworthy of analysis. To be honest, it is now extremely difficult to know when a distinctive sub-population is functionally absent from a survey sample. When 20% of the target sample used to be missing, the possibility of subgroups being completely absent was acknowledged but accepted. But when 50% of the target sample is missing, it is much more difficult, perhaps impossible, to make this claim. And if the obtained sample is devoid of *entire* sub-groups, its value may be limited even for exploring relationships.

## **Some Responses to these Challenges**

Because non-response is not a new challenge, survey administrators have developed and implemented a variety of measures in an attempt to maintain rates. These have attempted to address both refusals and failure to contact survey respondents.

Since refusals have been the major problem, most of the conventional measures are directed at reducing refusals, for example, offering incentives in the form of money (e.g., the Ontario Tobacco Survey) or a report of results (e.g., comparative SHAPES data for schools)<sup>5</sup>; sending letters to the household before phoning; and making an extra effort to convert refusals. While the more effective measures are those implemented as part of data collection, a few, such as post-stratification reweighting and imputation, can be useful after the data are gathered. But imputation is complex and

<sup>5</sup> See Brown (Appendix A).

not necessarily effective, while post-stratification weights can only be adjusted for Census-type characteristics of known distribution.

Another approach to improving response rates, perhaps more often suggested than tried, is to conduct fewer, better, surveys. One variation on this theme that has been used successfully is splitting the sample within a single survey in order to collect more data from the same number of persons e.g., CAMH Monitor in some years), thus perhaps reducing the overall number of surveys conducted. Overall, however, there has been a tendency over the past decade or two for more, and more specialized surveys, even though there may be significant commonality of content and target population (Stephens, Appendix A). Better communication among researchers at the planning stage may help reduce some of this overload on respondents.

While non-contact is less of a problem than refusals in cross-sectional surveys, it is a major issue with longitudinal designs. Some commonly used tactics to reduce non-contact, including loss-to-follow-up, are increasing the number of callbacks and scheduling calls more systematically in order to catch potential respondents at home. Placing more onus on call-bank providers to remove ineligible numbers may also improve the efficiency of initial call attempts.

In the opinion of some observers, these measures have only managed to slow the decline of response rates in recent years, not to improve them. Hence an increasing amount of effort is going into developing new approaches to these old challenges.

### **Innovative Approaches to Improving Response Rates**

Several types of initiatives in various states of development are currently underway.

Multi-mode surveys combine telephone interviewing with another mode of data collection, such as mail or web-based supplements (Thompson, Appendix A). While mail surveys are well established, they are unlikely to provide the boost in response rates that survey administrators seek. Web-based supplements show more promise, because of their inherent convenience for both collection and data processing. Developing methods to encourage timely and good quality responses to surveys on the web is a priority, as is developing and using web survey frames.

New sampling frames will be needed to pursue some of the developments that might increase survey response. For example, conventional RDD sampling frames will have to be adapted before cell phones are included in the banks used for surveys. While about 6% of Canadian households have only cell-phone service, many others have a cell phone and a traditional land line. When a household has multiple phone numbers, there is an increased probability of contacting it, requiring adjustment to sampling weights.

Expanded use of web-based surveys may also require new sampling frames. This will be a major challenge when such surveys are *not* a complement to an RDD or other household-based survey that has a frame associated with it. Some of this development work is being done on campus, where there is a known population with access to campus-based sites (Brown, Appendix A).

Most telephone sampling frames used by survey researchers are very limited in the information they provide, typically, a 10-digit working number and a geographic identifier that is only approximate. Nothing is normally known about the household until a call is placed, received, and a brief interview has been completed. When non-contacts constitute 50% or more of the called households, the result is not only absence of significant information, it also leads to a high level of inefficiency. This situation is exacerbated when the survey requires a targeted sample, such as smokers. Commercial survey houses have addressed this by creating panels of pre-screened survey participants (although the representativeness of such samples is unclear). But now, Statistics Canada is pursuing this type of initiative, creating a master sample with basic data on potential respondents. Making use of data from earlier survey contacts, this master sample will allow for more efficient targeting of potential contacts and may thus improve response rates by increasing the likelihood that householders will see the survey as personally relevant.

In this context of falling response rates, there is an important role for efficient screening of smokers and non-smokers at the first survey contact. Although not a solution to falling response rates *per se*, early screening of smoking status can help determine related bias in non-response, in cases where no interview is completed after the screening. The OTS has had some success in developing an efficient and simple screener question of this type.<sup>6</sup>

### **Innovative Approaches to Dealing with Non-response**

The working session briefly considered some measures that would address non-response at the post-collection stage. Such techniques are mostly variations on practices currently in place, and are not as innovative – or probably as promising – as those focused on improving response rates.

One suggestion involves post-stratification weighting of data, which amounts to a form of imputation by supplying missing data. The general principle is to use weights based on response patterns; an example is to weight Canadian Tobacco Use Monitoring Survey (CTUMS) data by in-school/not in school status. The objective is to create a more representative sample from data obtained from a sample deemed to be insufficiently representative. The challenge is to identify the

<sup>6</sup> Diemert LM, Bondy SJ, Victor JC, et al., Efficient screening of current smoking status in recruitment of smokers for population-based research. *Nicotine and Tobacco Research* 2008, 10:11,1663-1667

relevant parameters that affect representativeness (e.g., in-school status) and then adjust the sample accordingly, using data from an independent and credible source, such as the Census.

Another approach to the issue of small, unrepresentative, samples may be to combine data from separate, compatible surveys. While the opportunities for this will be limited, one example worth pursuing is to combine data from two ongoing surveys that include 15-19 year-olds, namely CTUMS and the Youth Smoking Surveys ([http://www.yss.uwaterloo.ca/ysssite\\_app/controller/index.cfm](http://www.yss.uwaterloo.ca/ysssite_app/controller/index.cfm)). There are a number of issues to consider in such a step, not the least of which is the appropriate weighting of cases according to in-school status, as noted above.

Regardless of the approach taken to dealing with missing cases, whether familiar—such as imputation—or novel—such as some of the weighting suggested here, sensitivity analyses should be performed to document the impact on survey results. While rarely performed, such analyses are informative and can be reassuring. An early example is the comparison of four sets of estimates for current smokers in the 1985 Health Promotion Survey: non-standardized (as is typical of most survey reports), age-sex standardized, age-sex-education standardized, and age-sex-income standardized. The standardized rates differed negligibly from the non-standardized rates, except when income was included. The effect of income was attributed to the high level of non-response for this variable.<sup>7</sup>

<sup>7</sup> Rootman I, Warren R, Stephens T, Peters L (eds). *Canada's Health Promotion Survey*. Ottawa: Health and Welfare Canada, 1988.

## Measurement Issues: A Secondary, but Enduring Challenge

As more and more surveys are conducted – and their results are compared - measurement issues will continue to merit attention. In a new handbook on the issue, the International Agency for Research on Cancer notes:

Consumers of survey data in which tobacco use measures are included should be aware of factors that can influence population estimates of tobacco use and take those into consideration when comparing estimates from surveys conducted within and across countries.<sup>8</sup>

### Cigarette Sales as a Measure of Consumption

The principal new measurement issue is the difficulty of estimating population consumption from cigarette *sales*. While the discrepancy between total self-reported consumption and cigarette sales data has long been recognized, new developments in tobacco retailing cast doubt on the reliability of sales data in their own right. This in turn undermines the value of per-capita consumption as a key indicator of progress in tobacco control.<sup>9</sup> Two developments were noted as underlying this measurement challenge (Kaiserman, Appendix A): (a) the increased amount of contraband and smuggled cigarettes being consumed but not reported, and (b) new practices by cigarette manufacturers whereby some of what they report as “wholesale” now includes distribution direct to the retail level. Solutions for both these challenges depend on regulation and law enforcement.

### Question Order and Context

A new example of an old issue was raised, illustrating the impact of different methodologies on population estimates of smoking prevalence (Snider & McAuley, Appendix A). In this case, the differing context and question order for probing smoking behaviour in the CCHS<sup>10</sup> and CTUMS<sup>11</sup> result in consistently higher estimates from the former survey. While the discrepancy is not major, and the temporal trends from both surveys are essentially identical, the difference raises questions as

<sup>8</sup> Measuring tobacco use behaviours. In IARC, *IARC Handbooks of Cancer Prevention: Tobacco Control. Volume 12. Methods for Evaluating Tobacco Control Policies*. Lyon, France: International Agency for Research on Cancer, 2009, Ch 3.1.

<sup>9</sup> Copley TT, Lovato C, O'Connor S. *Indicators for Monitoring Tobacco Control: A Resource for Decision-makers, Evaluators and Researchers*. Canadian Tobacco Control Research Initiative, 2006.

<sup>10</sup> Canadian Community Health Survey: <http://www.statcan.gc.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SDDS=3226&lang=en&db=imdb&adm=8&dis=2>

<sup>11</sup> Canadian Tobacco Use Monitoring Survey: [http://www.hc-sc.gc.ca/hl-vs/tobac-tabac/recherche-recherche/stat/ctums-esutc\\_2007-eng.php](http://www.hc-sc.gc.ca/hl-vs/tobac-tabac/recherche-recherche/stat/ctums-esutc_2007-eng.php)

to which estimate is the more accurate. Biochemical validation could shed some light on this, and will be available from Statistics Canada's new Canadian Health Measures Survey.<sup>12</sup>

## Measuring Youth Smoking

The working session recognized some measurement challenges that are specific to youth, including the low relevance of standard "adult" questions, need to focus on constructs such as initiation and propensity, and comprehension and reliability issues in reporting a complex behavior such as smoking when it is not an established routine. These issues were not discussed in detail; the suggestion was to consider (again) what variables should be monitored beyond tobacco use, e.g., psychological factors related to cessation, social support, and exposure to policy inputs. It is important to note that, with the institutionalization of the Youth Smoking Surveys, these issues are being addressed much more thoroughly than was the case in Canada a decade ago.

## Use of Standard Definitions

Two other measurement issues were placed on the record during the working session. Both are long-standing but apparently not regarded as very pressing, since substantive discussion was limited and solutions are easy, in principle: they involve only a change in well-established practices and some documentation to accompany the transition to the "new order." One of these issues is the failure to use standard definitions for key constructs across surveys.<sup>13</sup> The principal examples that remain 15 years after a federal workshop on this issue<sup>14</sup> are the 100-cigarette criterion for defining a smoker, and the absence of a specified duration of abstinence for defining a former smoker. Another issue, attributable more to the convenience of data gatherers than the needs of data users, is the lack of consistency in age groups for surveying youth, e.g., age 15-19 in CTUMS, age 12+ in CCHS, and age 18+ in the CAMH Monitor. In the absence of more uniform practice, discussion of this latter issue encouraged the reporting of youth data by single year of age where possible. This detail would not only permit ready comparison among surveys with different target populations, but it would also reveal the heterogeneity that exists within a five-year age group that spans developmental stages. These issues are well documented; their solution requires only a will to change established practices.

<sup>12</sup> CHMS: <http://www.statcan.gc.ca/bsolc/olc-cel/olc-cel?lang=eng&catno=82-003-S200700010362>

<sup>13</sup> An issue that is hardly unique to Canada, as documented by the IARC in their new Handbook (footnote 7).

<sup>14</sup> Mills C, Stephens T, Wilkins K. Data for Monitoring Tobacco Use. *Chronic Diseases in Canada* 1994; 15 (3): 105-110. Also in *Health Reports* 1994; 6 (3): 377-387.

## Next Steps: New Initiatives to Consider

Over the course of the discussion periods at the working session, there were several suggestions for new initiatives to address issues in telephone surveys of tobacco use. These can be roughly sorted into three categories – initiatives involving collaborative ventures, those requiring some R&D, and those requiring education/publication. None was discussed in sufficient detail to make the next steps obvious, however. More work will be needed.

### Collaborative Initiatives

Of the various initiatives identified, those involving collaboration are the most novel and may hold the greatest promise for meeting the emerging challenges. Indeed, one of these—to *create an ongoing working group to continue the discussion* – is virtually a precondition to achieving progress on the others:

- Establish an ongoing working group on response rate issues, including (or perhaps especially) the idea of a respondent panel to be used as a shared resource.
- Develop better coordination between federal and provincial players with respect to designing and conducting school surveys.
- Develop a national data base of school-level smoking rates to take advantage of natural experiments occurring within and among provinces.
- Find the will to implement standardized definitions and measurement procedures.

### Developmental Initiatives

Only one broad suggestion – expressed more as a hope than a substantive proposal – was advanced in response to the observation that the typical household survey systematically under-represents some groups in the population. That these groups are often heavy users of tobacco products adds urgency to this issue, but the very heterogeneity of the under-represented groups means that no single solution will be effective for all. Nevertheless, for the record, there was encouragement expressed to:

- Develop new methods for reaching special populations, i.e., currently under-represented subgroups.

### Educational/Publishing Initiatives

While writing and publishing a report is often seen as the easiest and most natural outcome of a working session, it was not in fact the first response after the discussion reported in this Special Report. Moreover, it was not suggested as an end in itself but as an important contribution to solving the problems discussed at the working session, at least with respect to response rates:

- Prepare reports (journal articles or other formats) on best practices for maintaining/increasing response rates in population surveys conducted by telephone.
- Prepare a commentary on how to describe response rates in survey reports and how to interpret them in microdata documentation.

## Conclusions

The conclusions most apparent after 10 presentations and 10 hours of discussion on the issues affecting telephone surveys of tobacco use are that:

- Household surveys by telephone are in trouble, due primarily to declining participation;
- More of the same tried-and-true remedies will not solve the problem;
- Wholly new approaches are needed.

In this spirit, the Working Session produced a large number of observations, suggestions, recommendations, and hopes that can be summarized under two headings – *Current Practices* and *Future Directions*.

### Current Practices

A variety of practices was identified as useful, especially for addressing response rate issues. Victor & Diemert (Appendix A) provided a nice classification of these for longitudinal surveys that also applies, in part, to cross-sectional surveys.

Perhaps the key point made regarding current practices is that, while they may have slowed the decline in response rates, they have not reversed them, nor will they. Since there is no prospect of returning to the high levels of response enjoyed 10-20 years ago, the focus was on the future and the need for new approaches.

### Future Directions

Many actual and potential next steps were identified during the presentations and discussions. All but a few were directed to the response rate issue; the few exceptions addressed measurement issues.

#### ***Survey Design and Planning***

- Conduct fewer, but smarter, surveys (more targeted, less repetition).
- Establish a coordinated ethics review process that is common across institutions and is consistent across the country.
- Adapt RDD sampling frames to include cell and internet phones.
- Develop sampling frames for conducting surveys on the web.
- Create pre-screened panels of potential survey participants.
- Develop, test and refine multi-mode surveys, especially web-based surveys to complement telephone surveys.
- Consider (again) what variables should be monitored beyond tobacco use.
- Use standard definitions for key constructs across surveys.

- Develop and implement a standard definition of “adults” in surveys.
- Develop better coordination between federal and provincial players with respect to designing and conducting general population and school surveys.
- Develop a national data base of school-level smoking rates.
- Establish an ongoing working group on response rate issues, with a focus on developing a respondent panel.

### ***Field Operations***

- Develop methods to encourage timely and accurate responses to surveys on the web.
- Place more onus on call-bank providers to remove ineligible numbers and to include cell and cable-based phones.
- Use biochemical validation to assess standard questions on smoking behaviour.
- Develop new methods for reaching populations that currently are under-represented.

### ***Research and Analysis***

- Identify the parameters that affect representativeness within existing survey data sets (e.g., in-school status) in order to adjust the sample through re-weighting.
- Conduct sensitivity analyses of the impact on survey estimates of various approaches to re-weighting and imputation.
- Look for opportunities to combine data from separate, compatible surveys.
- Conduct a thorough empirical examination of the factors behind declining response rates as a guide to developing effective solutions.
- Examine the association between survey context/question order and estimates of key indicators including smoking status.

### ***Knowledge Transfer***

- Document best practices for maintaining/increasing response rates and share this with survey practitioners and funding agencies.
- Prepare a commentary on how to interpret and describe response rates and share this with researchers, advocates, policy-makers, editors and funding agencies.
- Report data for youth (less than age 20) by single year of age where possible.

### ***Other Solutions***

- Use available regulatory and legal mechanisms to improve the quality of data on cigarette sales, by clearly distinguishing retail from wholesale sales, and containing contraband and smuggled products.

## **Appendix A: Speaker Summaries of Invited Presentations**

Appendix A: Speaker Summaries of Invited Presentations.....	18
What's Past is Prologue Even with Tobacco Use Surveys.....	19
Definitions and Changing Patterns of Response Rates .....	24
Measurement Issues in Tobacco Control.....	27
A Tobacco-Control Advocate Reflects on the Use of Survey Data .....	30
Representativeness of Federal Cross-Sectional (Tobacco) Surveys .....	33
Longitudinal Surveys and Attrition: Issues and Solutions .....	36
Methodological Issues in Reporting Cigarette Use and Consumption .....	40
Comparisons of School-Based and Telephone Surveys of Youth .....	45
Accounting for the Effects of Data Collection Method in Surveys Collecting Data Using Both the Web and the Telephone.....	51
Declining Response Rates: The Experience of the International Tobacco Control Policy Project .....	55

## **What's Past is Prologue Even with Tobacco Use Surveys**

Tom Stephens, Thomas Stephens & Associates

### ***Purpose and Scope***

Canada has a substantial history of surveying smoking behavior, with over four decades of monitoring the adult population. The data have been used to identify the need for tobacco-control policy, to assess progress in tobacco control, and, occasionally, to probe into the dynamics of maintenance and cessation. This presentation summarized the main features of surveys of adults and identified some enduring conceptual and methodological issues. The focus was on household surveys that target adults and are national in scope (Table 1). The history of youth smoking surveys in Canada is much shorter, and was not dealt with in this presentation.<sup>1</sup>

While Health Canada<sup>2</sup> initiated most of this data-gathering and published most of the reports, Statistics Canada also played a significant role in analysis as well as data collection. Surveys of smoking have been both special purpose (primarily those by Health Canada) and more broadly focused on health (primarily those by Statistics Canada). Tobacco content has always concentrated on behavior, with only occasional attempts to probe knowledge, attitudes, and intentions; more recently, exposure to ETS has become a fairly routine topic (Table 1).

The history of this survey enterprise has been impressive in its duration and the data have been widely and well-exploited. However, it has also been marked by inconsistent definitions and irregular intervals, until the advent of true time series with CTUMS and CCHS. The longest-running and most frequent series is not by government but by the tobacco industry (Table 1). However, its methods and definitions are not publicly accessible. The longest running reputable time series in Canada is actually not national in coverage; it is the CAMH Monitor and its predecessors.<sup>3</sup>

### ***Looking Back: Issues for Secondary Analysis***

A number of issues, primarily of a definitional nature, can bedevil secondary analysis of old data sets, particularly when this involves comparison among different surveys. Often, these issues are not apparent from the survey questionnaires but require close reading of survey documentation.

### **100 Lifetime Cigarettes**

Despite recommendations on this issue from a Health Canada workshop in 1994<sup>4</sup> and the fact that most surveys ask if the smoker has consumed a lifetime total of at least 100 cigarettes, this variable is

<sup>1</sup> For background on the Youth Smoking Surveys, see [http://www.yss.uwaterloo.ca/ysssite\\_app/controller/index.cfm](http://www.yss.uwaterloo.ca/ysssite_app/controller/index.cfm)

<sup>2</sup> In the early days, "Health and Welfare Canada".

<sup>3</sup> [http://www.camh.net/research/camh\\_monitor.html](http://www.camh.net/research/camh_monitor.html)

<sup>4</sup> Mills C, Stephens T, Wilkins K. Data for Monitoring Tobacco Use. *Chronic Diseases in Canada* 1994; 15 (3): 105-110. Reprinted in *Health Reports* 1994; 6 (3): 377-387.

not routinely used to define current or never smokers. The *CAMH Monitor* is the sole significant exception among the surveys in Table 1.

### **Distinguishing Never from Former Smokers**

Through 1974, the Labour Force Survey supplements combined never and former smokers and reported them simply as non-smokers.

### **Gender-specific Definitions**

Until 1974, male and female non-smokers were defined differently. Cigar and pipe use was taken into account for males but only cigarettes counted for females.

### **Proxy Data**

Third-party reports of smoking behaviour were accepted by the Labour Force Survey through 1981. When it became possible to distinguish proxy from self-reports, it became clear that the proxy reports for teens seriously underestimated their smoking.<sup>5</sup>

### **Age Groups**

Perhaps because of the convenience of attaching the earliest smoking questions to a survey of labour force activity, coverage began at age 15 and this practice remains the norm, with the target population being typically described as “adult.” Not only are 15-17-year olds not adult by any definition, but including them complicates comparisons with true adult surveys that start at age 18 (e.g., the *CAMH Monitor* and many US surveys). When data are not available by single years of age, it can be impossible to compare the youngest age groups, and the total of all ages, among different surveys. Moreover, the age group 15-19 years is so heterogeneous with respect to smoking that it makes little sense to publish a single prevalence number for them (see Brown, Appendix A).

### **Looking Ahead: Enduring Issues for Future Surveys**

Apart from banishing proxy data and correcting some definitional quirks, some of these old practices remain current. They should be seriously addressed whenever a new survey is being considered. The more important include setting an appropriate lower age for adult surveys (15 or 18 years), and using the 100-cigarette criterion to define smoking status. The other issue that is significant in Table 1 is the diminishing representativeness over time of longitudinal samples. This proved to be a major theme in the working session.<sup>6</sup>

<sup>5</sup> Stephens T. *A Critical Review of Canadian Survey Data on Tobacco Use, Attitudes and Knowledge*. Prepared for Tobacco Programs Unit, Health and Welfare Canada, 1988.

<sup>6</sup> See Victor & Diemert, Thompson, and Boudreau in this Appendix.

**Table 1: National Surveys of the Adult Household Population with Significant Tobacco Use Content, Canada, 1965 – present**

Year(s)	Survey (age coverage)	Sponsor/ analysis	Overall focus of survey	Tobacco content	Comments
1965, 1966, 1970, 1972 1974	<i>Labour Force Survey [LFS] supplement (age 15+)</i>	Health Canada	Cigarette smoking	Smoking status, # cig. smoked	Early years accepted proxy reports → questionable teen data. Never/former not distinguished before 1975. Very large sample.
1975, 1977 1979			Cigarette smoking (plus cigars and pipes)	Smoking status, # cig. smoked, usual cigarette brand, cigar/pipe use	Content describes 1979 version. Brand and amount → tar/nicotine exposure.
1976 - pres, annually	ITL Incidence and Usage (18+?)	Imperial Tobacco Limited	Cigarette smoking	“Incidence” and “usage” i.e., prevalence and 5+ cig. daily	Data courtesy David Sweanor. Methods undocumented. More consistent surveying than by any government is notable.
1977, 1982, 1984, 1987, 1989, 1991-pres, annually	CAMH Monitor and predecessors [Ontario Drug Monitor, Ontario Alcohol and Other Drug Opinion Survey, Adult Drug Use] (age 18+, Ontario)	ARF/CAMH	Substance use, gambling, mental health	Smoking status, # cig. smoked, SHS exposure, other behaviours, policy attitudes	OK, it’s not national, but the long, consistent history deserves mention. Core data have been consistent while special topics have varied over time. First use of 100-cig. criterion?
1978-79	Canada Health Survey [CHS] (age 15+ for risks)	Health Canada & Statistics Canada	Risk factors, health status, use of services	Smoking status, # cig. smoked, usual brand, inhalation, quit attempts, cigar/pipe/ cigarillos	Intended to be ongoing, but cut prematurely by HC. Grandpa of CCHS, stepfather of Canadian Health Measurement Survey
1981 1983 2002-4	Canada Fitness Survey [CFS] (age 10+, Wave 1)	Canadian Fitness & Lifestyle Research Institute	Physical fitness and activity	Smoking status, cigar/pipe/cigarillos, spouse’s smoking	Longitudinal design, non-standard tobacco questions
1981 1983	Labour Force Survey [LFS] supplement (age 15+)	Health Canada	Cigarette smoking (plus cigars and pipes)	Smoking status, # cig. smoked, age started, brand switching, inhalation, quit attempts, when stopped	Brand not specified, just switching and filter use.
1985	Health Promotion Survey I [HPS] (age 15+)	Health Canada	Risk factors: behaviour, beliefs, attitudes; health	Smoking status (limited), avoiding SHS, opinions and beliefs	Never/former not distinguishable. First publication of smoking prevalence x province (not region)

## The Future of Household Telephone Surveys on Tobacco: Methodological and Contextual Issues

Year(s)	Survey (age coverage)	Sponsor/analysis	Overall focus of survey	Tobacco content	Comments
			status		
1985	General Social Survey I [GSS] (age 15+)	Statistics Canada	Health status, use of services, health behaviors, social support	Smoking status, # cig. smoked, brand, cigar/pipe/cigarillos, household smokers, age started/stopped	
1986	Labour Force Survey [LFS] Supplement (age 15+)	Health Canada	Cigarette smoking (plus snuff and chewing tobacco)	Smoking status, # cig. smoked, brand, workplace smoking restrictions, attitudes toward restrictions	Brand asked but not published. Cigar and pipe use dropped.
1989	National Alcohol and Drugs Survey [NADS] (age 15+)	Health Canada	Drinking, use of other drugs	Smoking status, # cig. smoked, age started, year stopped	
1990	Health Promotion Survey 2 [HPS] (age 15+)	Health Canada	Risk factors, social support	Smoking status, # cig. smoked, restrictions on smoking, household smokers; quit attempts, intentions, aids	
1991	General Social Survey 6 [GSS] (age 15+)	Statistics Canada	Health status, use of services, health behaviors	Smoking status, # cig. smoked, brand, household smokers, age started/stopped	
1994	Canada's Alcohol and Other Drugs Survey [CADS] (age 15+)	Health Canada	Substance use, especially alcohol	Smoking status, # cig. smoked, age started, year stopped, quit attempts,	First use of 7-day wheel for amount smoked, borrowed from alcohol use surveys.
1994-5 to present, every 2 yr	National Population Health Survey [NPHS] (age 12+, Wave 1)	Statistics Canada	Risk factors, health status, use of services	Smoking status, # cig. smoked, smoking during pregnancy, nicotine dependence, contemplating quitting, beliefs about smoking (1996)	Longitudinal design → diminishing representativeness.

Year(s)	Survey (age coverage)	Sponsor/analysis	Overall focus of survey	Tobacco content	Comments
1994-95	Survey of Smoking in Canada [SOSIC] (age 15+, Wave 1)	Health Canada	Cigarette smoking	Smoking status, # cig. smoked, recent changes, impact of 1994 price change	Longitudinal design: 4 cycles in 12 months → bias through severe attrition.
1999 – present, semi-annually	Canadian Tobacco Use Monitoring Survey [CTUMS] (age 15+)	Health Canada	Cigarette smoking	Smoking status, # cig. smoked, use of light/mild and reasons, brand, source, smoking and restrictions at home, age started, when stopped, quitting history, cessation aids, smoking and pregnancy, SHS exposure, attitudes, use of cigars/pipe/cigarillos/snuff/chew	Core data have been consistent while special topics have varied over time. 100-cigarette lifetime question is asked but not used in HC analysis to define current/daily/occasional smoker. It is used to define former smokers.
2001–2005 every 2 yr, 2007 - present, annually	Canadian Community Health Survey [CCHS] (age 12+)	Statistics Canada	Health status, use of health services, risk factors	Core: Smoking status, # cig. smoked, source of cigarettes, smoking and restrictions at home, age started, when stopped, cigars, pipe, snuff, chewing tobacco, SHS exposure. Optional: quit aids	Very large sample. Consistent core content plus optional modules per health region. 100-cig. lifetime question asked but not typically used in STC analysis.

## **Definitions and Changing Patterns of Response Rates**

David Northrup, Institute for Social Research, York University

Practitioners and users of survey data have always been concerned about response rates. While there is variation by country, by academic discipline, by area of public policy, by mode of data collection, and by survey house, there has been a general pattern of periods of heightened concerns about declining response rates followed by improvements in data collection methods. By and large, forty years ago, when reputable survey firms were getting response rates in the 70 to 80 percent range, and central statistical agencies were getting response rates of 90 percent or better, high response rates were seen as evidence of a ‘good survey.’ Today, when well designed, reasonably funded surveys about public health issues have rates in the 50 percent range, or lower, definitions of what is a ‘good survey’ have changed.

Commonly accepted definitions of response rates date back to the early 1980s. Steeh, writing in 1981 in what has become a touchstone article in *Public Opinion Quarterly*, noted “increasing concern over response rates in sample surveys” in the previous decade. While documenting declines in response rates, Steeh also reported that awareness of these declines has led to “experimentation with interviewing techniques.” Given that by 2006 leading figures in survey research were questioning the relationship between response rates and sample bias, she also was prescient in her view that “greater numbers of non-respondents create problems only when they are systematically distributed with respect to major survey variables.”

Following Steeh in 1982, the Council of American Survey Research Organizations (CASRO) offered a standard definition for a response rate that was generally accepted by state statistical agencies, academic survey centres, and commercial firms. In simple terms, response rate was defined as: completed interviews / an estimate of the number of eligible sample units. The definition took into account the increasing number of non-respondents that resulted from non-contacts. In household telephone surveys, this meant estimating the number of households hidden in the ‘never answered’ or ‘always busy’ numbers where an interview could have been completed, if at least one of the interviewers’ calls had been answered.

The easiest way to boost a response rate, by reducing the non-contacts, was to increase visits to prospective households or make more telephone calls. In the mid-to-late 1980s, with the exception of state statistical agencies, most population-based public health surveys moved from in-home data collection to telephone data collection. As researchers migrated to the telephone, the number of attempts to reach respondents increased dramatically -- no doubt because adding more telephone calls before giving up on a potential interview was much cheaper than revisiting homes. During this time period, efforts were also put into determining the best possible sequence of call attempts to obtain an interview by using computers to track every call attempt to every possible household.

While these efforts may have slowed down the rate of decline in response rates, increasingly the major factor in declining response rates resulted from refusals to participate by the individual or

household. In an effort to combat increasing refusals, increasing effort was put into understanding the relationship between interviewer characteristics (particularly interviewers' voices) and response rates as well as the affect of interviewer training on response rates. Greater attention was paid to survey introductions and, as reverse telephone directories have become available, the use of advance letters. Increasingly, with hard-to-reach populations, cash incentives have been used.

By the mid-1990s, most survey organizations were again reporting sharp declines in response rates. The Behavioral Risk Factor Surveillance Survey (BRFSS),<sup>1</sup> centered at the Centre for Disease Control (CDC), is a large public health survey that collects data on a number of key health issues, including smoking behaviours. BRFSS reported declines in response rates from the low 60s in 1996 to the low 50s by the year 2000. Closer to home, the response rate for the Centre for Addiction and Mental Health's CAMH Monitor<sup>2</sup> follows the same trend. There was a steady decline in response rates from 69 percent in 1996 to a low of 55 percent in 2008.

Between 1996 and 2008 there has been a considerable effort to ameliorate the decline in the CAMH response rate. In 1996, the average number of call attempts required to obtain an interview was 4.14 and by 2007 it was 5.43. The number of interviews completed on the tenth or subsequent call attempt increased from nine percent to sixteen percent for the same time period. These efforts have helped to minimize non-response resulting from non-contact. Only about 20 percent of the total non-response on CAMH Monitor is a result of non-contacts. The remaining 80 percent is accounted for by individual and household refusals. In 1996, the number of 'converted refusals,' as a proportion of the total number of interviews, was 4.5%; by 2007 this figure had almost quadrupled to 17.7%. Despite these efforts, the response rate for the CAMH Monitor has declined by 14 percentage points over the last 12 years.

When it comes to response rates, the prognosis for the future is not encouraging. There is no reason to think response rates will not continue to decline for a number of years. More worrisome is the sharp drop in response rates for Institute for Social Research (ISR) studies in the last two years, a trend also noted for some American studies, including the BRFSS. New data collection strategies, such as the use of multiple modes of data collection (for example, using mail and web surveys to supplement telephone surveys), and new sample strategies, such as the use of panels, may mitigate these declines in response rates. A reading of the history of survey research, however, suggests these efforts will slow down, or at best temporarily stall, declines in response rates, not reverse the trend.

The only potential silver lining to this increasingly dark non-response cloud is the accumulation of research that demonstrates that response rates are often not correlated with sample bias. Public Opinion Quarterly, arguably the world's leading journal with respect to survey research methods for the social sciences, dedicated a Special Issue to Nonresponse Bias in Household

<sup>1</sup> <http://www.cdc.gov/BRFSS/>

<sup>2</sup> [http://www.camh.net/research/camh\\_monitor.html](http://www.camh.net/research/camh_monitor.html)

Surveys in 2006. In the lead article, Robert Groves completes a meta-analysis of 30 reports on the relationship between response rate and sample bias. His general conclusion is that “nonresponse biases in estimates are only indirectly related to nonresponse rates.”

As Steeh suggested 25 years earlier, response rates are only found to affect estimates when the likelihood of responding is related to variables central to our analysis. While this conclusion is comforting at a general level, it leaves many open questions for those of us who conduct surveys about public policy with respect to smoking and surveys with smokers about their behaviours.

What, for example, is the correlation between willingness to complete a telephone survey and support for restrictions on smoking in public places? Or perhaps more importantly, when we complete surveys only with smokers, is agreeing to participate correlated with values or smoking behaviours we want to measure? What we need to do is better understand the relationship between willingness to participate and attitudes, values and behaviours of smokers. This is a tall order, but there are a number of places to start.

First, we need to better understand the relationship between current efforts to increase response rates and sample estimates. In terms of making many call attempts to obtain an interview, we need to know if ‘hard-to-reach’ smokers are different than ‘easy-to-reach’ smokers. Are estimates for behaviours for smokers who are ‘converted refusers’ the same as that for standard interviews? If multiple modes are used to collect the data, is there a relationship between mode of data collection and estimates? For what estimates can data external to the survey be used to evaluate our estimates? When we are measuring change in estimates, do repeated cross sections (the choice for most public health surveys) or longitudinal designs produce better data?

Answering these questions will require bigger, more complex, and more costly surveys. There is an issue of where funds will be found. As a provider of survey data, and at the risk of speaking against my class interest, my advice for those of you whose research careers are based upon conducting surveys and analyzing the data, is that you should conduct fewer, but better surveys. Every survey you conduct should be seen as an opportunity to better understand the relationship between participation and estimates for your key variables of interest.

## **References**

Council of American Survey Research Organizations. *Special Report: On the Definition of Response Rates*. Port Jefferson, NY, 1982.

Groves, Robert M. Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly* 2006, 70(5): 646-75.

Steeh, Charlotte G. Trends in Nonresponse Rates, 1952-1979. *Public Opinion Quarterly* 1981, 45(1): 40-57.

## **Measurement Issues in Tobacco Control**

Sue Bondy, Dalla Lana School of Public Health, University of Toronto

### ***Introduction***

At this workshop, two major problem were discussed with respect to the utility of survey data – one being declining response rates (willingness of people contacted for surveys to take part), and the second (potentially even more insidious concern) being poor coverage or systematic exclusions from sampling frames. The present talk sought to clarify the methodological implications of loss of representativeness.

The real risk of poor sample representativeness is the potential for bias in survey estimates. What we estimate using survey data may be entirely descriptive statistics (proportions and means, for example), but may also include measures of association with related hypothesis tests. The risk of non-representativeness is that our estimates may not be externally valid, or not approximate the true population parameter for these values being estimated.

### ***Two Types of Survey User***

This talk identified important differences between uses (and users) of survey data in terms of what was being estimated -- because not all survey data users will be equally affected by response rates. With homage to Robert Groves (Groves, 1989), two hypothetically distinct users of survey data were introduced, namely: the “describer”; and, the “modeller”. The describer might be a professional in a public health or government setting responsible for monitoring and surveillance. This person’s paramount concern is the ability to generalize estimates of interest (typically descriptive statistics, such as rates and proportions) to a defined population.

It is typical, however, that one’s population (defined by political boundaries) is in fact a collection of sub-groups which differ in the prevalence of health states and behaviours under study. For example, smoking rates vary by demographic characteristics which, in turn, are associated with exclusion, contact and refusal patterns. Describers (such as public health epidemiologists) then, tend to be critically aware of when missing observations are non-ignorable concern and bias corrections need to be applied. However, they may be stymied, by budget and technical restraints, to do much about it. Probability and post-stratification weights are fairly easy to develop and a familiar tool for most analysts; but they are feeble measures against selection bias with response rates at 50% or less; and do not directly address frame exclusions.

Solutions may include using multiple whole protocols for surveillance which combine sampling frames and modes of data collection. Examples of this, of course, can already be found in the Canadian public health scene. The Canadian Community Health Surveys include separate ‘household’ and ‘institutional’ sections, and studies of homeless populations have been carried out. Public Health indicators for surveillance often include multiple measures of the same underlying phenomenon from different sources, each or all of which having their own strengths

and weaknesses (e.g., prevalence of mammographic screening, or diabetes prevalence, from self-report data as well as from health care records). Using multiple surveillance methods multiplies costs and effort. It also creates difficulty (and political strife) in choosing between estimates which disagree - and weighted recombination of estimates from multiple sources is a challenge to develop methodologically (see Korn & Graubard, 1999), as well as to defend politically.

Finally, hard-to-reach populations are, well, hard-to-reach for many reasons in terms of politics and data collection feasibility. They may also be hard to identify (without adequate data to know where health experience is distinctive). Once identified, it may not be possible to launch custom-designed studies for multiple special populations in a way that permits combination with more mainstream survey cycles. Overall, the describer is severely affected by losses of representativeness to telephone surveys, and is most in need of creative, often expensive and onerous, means to fill gaps and correct for selection biases.

### ***The Modeller***

In contrast to the describer, the modeller testing hypotheses and building theory may be in a position to forego representative sampling altogether. It is worth remembering that many experimental studies (e.g., clinical trials or laboratory-type studies) are routinely done using highly restricted, volunteer and convenience samples. This is acceptable because internal validity is the first concern, and because external validity refers to application of findings to a singular clinical sub-population or defined set of conditions (as distinct from any population-averaged net benefit). Modellers are interested in testing hypotheses about putatively causal associations (would-be experimentalists, if you like) and estimating the likely impact of public health measures intended to effect change.

To the modeller collecting data from scratch, deliberately non-representative samples are often by far the most cost-effective (Groves, 1989). When modellers use existing survey data to test hypotheses, they develop multivariable models to control for confounding and bias (Szklo et al., 2000). These models may not reflect averaged effects seen in the underlying population (Korn & Graubard, 1999; Szklo et al., 2000).

Poor response rates (even probability weights) may be completely ignorable in analytic research (Groves, 1989; Korn & Graubard, 1999). To see this, one has to understand how estimates of association (correlation coefficients, odds ratios, and the like) are affected by non-representativeness. For unequal selection probabilities to affect an estimate of association, that association has to vary within the population (Groves, 1989; Korn & Graubard, 1999). There has to be some form of interaction (or effect modification) *and in turn* the strength or direction of association has to be correlated with selection probability. If one is studying the benefit of a measure to encourage smoking cessation, it isn't sufficient that non-respondents differ in smoking prevalence. It must also be true that the magnitude of intervention benefit varies (on average) between non-respondents and respondents. Measures of association are more robust to non-representativeness than descriptive statistics (Groves, 1989). The analysts may deal with

them by restriction to one sub-group of interest at a time or by modelling interaction effects, as opposed to population averaged effects (Korn & Graubard, 1999; Szklo et al., 2000).

### **Common Interests**

Modellers too, though, may share the interest in describing population-meaningful effect estimates (particularly in policy evaluation). Certainly, modellers and describers (we are both) have interests in common in seeing that large multi-purpose health surveys receive funding and political support. Large samples needed for finely detailed descriptive surveillance may not be justifiable for this purpose alone. The use and re-use of data (by researchers and students in a range of settings across the country) increases the public profile and benefits of investments made into surveillance research. Users contribute directly when paying for data access. Similarly, many public health problems lack the political profile such that surveillance research would be carried out (at all or often) unless multipurpose studies with many investigators were already underway.

As response rates fall, there is a risk that (and reasons why) analytically-focused researchers may turn away from multipurpose surveys. Researchers may benefit from using registries of people willing to take part in research (saving recruitment costs). Similarly, peer-reviewers are trained to be critical of surveys with poor response rates – regardless of how articulate the author is in explaining why it doesn't invalidate their conclusions– so one might be better off with a research design where 'survey' isn't in the title to set up this expectation. Finally, hypothesis-testing researchers are, and should be, motivated to use stronger designs than cross-sectional research. This implies investment in prospective and quasi-experimental designs, rather than efforts to recruit reluctant populations.

For both modellers and describers, decreased reliance on telephone surveys is going to mean a lot of change in how we work. One hopes it doesn't lead to increased separation of researchers interested in the population as a whole from those investigating putative causal mechanisms and developing theory. Particularly on the side of surveillance and monitoring, there are real challenges ahead in terms of how to document health burden and trends. Research methods are going to have to be adapted, and a great deal of technical expertise will have to be developed to reach our objectives.

### **References**

Groves, Robert M. 1989. *Survey Errors and Survey Costs*. New York: Wiley.

Korn, Edward L. and Graubard, Barry I. 1999. *Analysis of Health Surveys*. New York: Wiley.

Szklo M, Nieto FJ, Miller D. 2000. *Epidemiology: beyond the basics*. Gaithersburg, Maryland: Aspen Publishers.

## **A Tobacco-Control Advocate Reflects on the Use of Survey Data**

Diane Kunyk, University of Alberta

Not long ago, Canadians could smoke virtually anywhere they pleased—at work, on public transport, in classrooms and even in hospitals. In the 1960s, our country had one of the highest rates of smoking in the world. Tobacco control measures have had substantial impact in Canada; and a great share of that success can be attributed to advocacy efforts.

Advocacy is considered one of the pillars of tobacco control. Within all six of the key tobacco control policies in the WHO's MPOWER (2008) package, which builds on the Framework Convention on Tobacco Control, a role for advocacy can be carved out. In their work to prevent the diseases and early death related to the use of tobacco products, some health care providers insist that advocacy is the sixth 'A' of tobacco intervention following "ask, advise, assess, assist and arrange". There are health care professional bodies that agree; in Canada, both the Royal Colleges of Physicians and the Canadian Nurses Association include advocacy as a role for their membership. Many researchers are also advocates, for example leading change in cigarette packaging and policies for smoke-free vehicles.

Survey data has a crucial role in responsible advocacy. To be effective, policy formulation must be based on current and reliable information (Pal, 2006). The public and decision-makers welcome empirical findings that demonstrate the potential to improve health, and the readiness to do so (Lomas, 1997). Almost routinely, the first questions asked regarding policy issues are: how many people are affected, what are others doing in this regard, what are the associated costs to change, and where do most people draw the line where the risk to health is worth restricting some freedoms? These are all questions that can be answered by survey data.

High-quality data is critical for effective advocacy work. We know that any assertion for further tobacco control endeavors will likely be met with resistance from the tobacco industry, affiliated industries profiting from tobacco use and sponsorships, and political parties grateful for their donations. Inaccuracies in reporting data quickly erode any credibility for the advocacy endeavor. In tobacco control, we know that the data and proposed strategy are right when they elicit a positive 'scream test' from the industry.

In return, advocacy can also advance the integration of science into practice (Chapman, 2007). Many potentially newsworthy scientific articles are read only by academics when accessed through library searches. Public health advocates may be able to focus the attention of the media as well as policy-makers on new and emerging evidence.

In Canada, we have a rich body of survey data when compared to many other countries. These data help us to make our case. For example, when pointing out the risks to youth associated with the use of little cigars, it is invaluable to note that their sales have increased over the past five years from fewer than 50,000 units in 2001 to more than 80 million in 2006 (Health Canada, 2008). When this is reported along with the finding that one-third (37%) of youth aged 15 years and older have ever tried little cigars, and 4% reported smoking little cigars in the past 30 days (CTUMS, 2007), it draws attention to the immediacy of this policy problem.

One universal advocacy goal is to minimize gaps between what is being implemented and what has been demonstrated empirically to make a difference. The reported increase in youth smoking in Alberta from 15% in 2006 to 20% in 2007 (CTUMS, 2007) provides an opportunity to critically appraise the efficacy of existing programming. Our voiced concern about the inability of Alberta physicians to receive remuneration for treating tobacco dependence, and the lack of cessation products on the provincial formulary, is made more meaningful when supported by survey evidence indicating that Alberta is an outlier when compared to other provinces and territories. When preparing our policy arguments for motions to restrict tobacco industry funding in some faculties at the University of Alberta, Joanna Cohen's (2001) work on academia and policy was invaluable.

There are those in the community who are not in a position to advocate and they must be the focus of our particular attention. Simon Chapman (2007) observed that "to smoke today in many nations is to wear a badge that says, 'I am either an immature youth, have little education or life aspiration, or I am a resigned addict.'" I would add the words 'or have a mental illness' to his line. It is critical that the needs of our marginalized populations are heard, and that we remove the barriers to the policy development that these groups require.

Unfortunately, the fine print of many of our surveys notes that sampling did not occur in prisons, on reserves, in mental institutions or in homeless shelters. Many of our youth, working poor, men in the trades, and transient populations, are also excluded from our surveys because if they have a phone, it is a cellular one. And these are some of the groups that are most at risk for smoking.

There is also evidence emerging to suggest that some of our most vulnerable members of society are not as responsive to existing tobacco reduction strategies. As a result, they may become further marginalized because of their high rates of smoking. The tobacco-control community is morally compelled to decrease these disparities as we have unintentionally increased the stigma they experience. We need high-quality survey data in order to advocate for informed policies and programming to meet their needs.

Lomas (1997) correctly points out that ‘Research is not a retail store’. Just because advocates require these data does not mean they can be made readily available. There are numerous methodological hurdles to reaching these marginalized populations. Not only do these raise the complexities of doing research but they likely also raise the costs.

This advocate concludes that including our vulnerable and marginalized populations in our tobacco control surveys is one of the challenges that needs to be addressed in the 21<sup>st</sup> century.

### **References**

Chapman, S. (2007). *Public health advocacy and tobacco control: making smoking history*. Oxford: Blackwell Publishing.

Cohen, J. (2001). Universities and tobacco money. *BMJ* 323:1-2.

Health Canada (2007). *Canadian Tobacco Use Monitoring Survey (CTUMS): Annual Results*. [http://www.hc-sc.gc.ca/hl-vs/tobac-tabac/research-recherche/stat/ctums-esutc\\_2007-eng.php](http://www.hc-sc.gc.ca/hl-vs/tobac-tabac/research-recherche/stat/ctums-esutc_2007-eng.php)

Health Canada (2008). A proposal to regulate little cigars under the Tobacco (Access) Regulations. [http://www.hc-sc.gc.ca/hl-vs/tobac-tabac/commun/consultation/\\_cigar/index\\_e.html](http://www.hc-sc.gc.ca/hl-vs/tobac-tabac/commun/consultation/_cigar/index_e.html)

Lomas, J. (1997). *Improving research dissemination and uptake in the health sector: beyond the sound of one hand clapping*. McMaster University Centre for Health Economics and Policy Analysis: Policy Commentary. C97-1.

Pal, L. (2006). *Beyond policy analysis: public issue management in turbulent times*. 3<sup>rd</sup> ed. Toronto: Nelson.

World Health Organization (2008). *WHO report on the global tobacco epidemic: The MPOWER package*. Geneva: WHO.

## **Representativeness of Federal Cross-Sectional (Tobacco) Surveys**

Judy Snider, Tobacco Control Directorate, Health Canada and Julie McAuley, Health Statistics Division, Statistics Canada

### **Overview**

Statistics Canada, in partnership with Health Canada, conducts two national cross-sectional surveys on adults with core tobacco content. These are the Canadian Tobacco Use Monitoring Survey (CTUMS) and the Canadian Community Health Survey (CCHS).

The CTUMS, which has been conducted since 1999, was developed to provide Health Canada and its partners with timely, reliable, and continual data on tobacco use and related issues. The survey's primary objective is to track changes in smoking status and the amount smoked, especially for 15–24-year-olds, who are most at risk for taking up smoking. It provides national and provincial estimates for half-year (waves) and full-year (annual). While the core questions of CTUMS do not change, the non-core questions have been and can be changed every six months to address key issues related to tobacco control as they emerge.

The CCHS, which has been conducted since 2001, has two components – the annual survey (formerly known as the x.1 cycles) and focused content survey (formerly known as the x.2 cycles). The presentation will focus solely on the annual survey, the primary objectives of which are to provide timely, reliable, cross-sectional estimates of health determinants, health status and health system utilization across Canada. Data are made available at the sub-provincial levels of geography, predominantly at the Health Region level. The questionnaire is comprised of core, themed and optional content, as well as a rapid response component.

There are a number of factors which may impact the representativeness of cross-sectional surveys regardless of the topic and coverage. The presentation discussed methodological factors including differences in survey design, new technologies, sub-populations and potential biases. Other survey methods were also described.

### **Factors Affecting the Representativeness of Cross-sectional Surveys**

There are important methodological differences between the CTUMS and CCHS. These include differences in the target population and sample sizes, the number of respondents per household, the frames used for sampling, the principal topics of the survey, and where in the questionnaire the question on smoking behaviour is asked. Regardless of these differences, we see that the decreasing trend in smoking rates for persons 15 years of age and older remains across both surveys, but the rate from CCHS is consistently higher (but declining). The differences, which

range from 2.5% to 5%, may be attributed to the differences in methodology between the two surveys.

However, there are a variety of challenges encountered when collecting data from respondents. These are highlighted in the presentation and are designed to provide interested individuals with a framework on the best survey design to ensure overall representativeness of their data.

### ***Cellular Telephones***

Fewer than one-quarter of Canadian households now use only traditional land-line telephone service as the nation shifts more and more to cell phones, cable or "Voice over Internet Protocol" telephone services (Residential Telephone Services Survey (RTSS 2007)). In December 2007, 6.4% of households reported they had only a cell phone, up from 5.1% a year earlier and 1.9% in 2003. This is a challenge for surveys, especially those such as CTUMS that use Random Digit Dialing (RDD) as a sampling method, as cellular phone have traditionally not been included in the telephone number data banks. Based on Statistics Canada's RTSS and a study undertaken by Blumberg of the National Center for Health Statistics, we know that certain populations are more likely to have only a cellular phone. These include low-income households, households in large urban centers, households in Western Canada, renters and people living with roommates, people going to school, and younger people (aged 15-24), particularly young males.

### ***Hard-to-Reach Populations***

Hard-to-reach populations, including newer immigrants, the homeless, students, the mentally ill, and the institutionalized, are important groups for representativeness of survey results. However, the counts within each of these populations are small. Reaching these populations is quite costly given that they are hard to reach and are typically under-covered.

### ***Validity Issues***

Regardless of the telephone survey, the interviewer has to assume the respondent is being truthful through the screening stage. As for response bias, issues that arise include respondents opting out of the survey by responding that they do not smoke when in fact they do, as well as biased recall of the amount they smoke and their exposure to second-hand smoke.

### ***Alternate Collection Modes***

While telephone and personal interview surveys are typically conducted by Statistics Canada, many other collection modes are available. Internet surveys are becoming more prevalent; however, their representativeness is tied to the penetration of internet access in Canadian households, which is known to be correlated with socio-economic status, literacy levels, and age.

Intercept surveys (e.g., stopping someone on the street, or a mall, etc.) allow for face-to-face interviews but suffer from selection bias (convenience sample, location of intercepts, willingness to participate on site or at a convenient place in the not too distant future). Mail surveys typically suffer from high levels of non-response.

### **Conclusions**

There are many challenges when conducting, analyzing and interpreting results from surveys that measure tobacco use in Canada. Nevertheless, the independent results from these surveys reveal similar downward trends in tobacco use over time.

The need to use more than one metric to measure smoking behaviour in Canada is becoming more evident. Data from the Canadian Health Measures Survey (CHMS) will be a powerful tool to help validate results and will allow for the examination of possible bias (amount and direction) in cross-sectional surveys by providing self-report and complementary bio-monitoring (e.g. cotinine, lead) data for smoking behaviours.

Although hard-to-reach populations have been identified as important to the representativeness of surveys, over time their impact on smoking prevalence will grow as they are not only hard to reach for surveillance, but also for prevention and cessation programs. Efforts will need to be made to reach these populations to ensure they are targeted in tobacco control efforts.

### **References**

Blumberg S & Luke J. Wireless Substitution: 2005 Update and Its Impact on Surveys of Children. *American Association for Public Opinion Research, 61<sup>st</sup> Annual Conference*, Montreal, Quebec, May 2006.

Statistics Canada. Residential Telephone Services Survey, December 2007. <http://www.statcan.gc.ca/daily-quotidien/080423/dq080423d-eng.htm>

Statistics Canada. Residential Telephone Services Survey, 2003. <http://www.statcan.gc.ca/daily-quotidien/060405/dq060405b-eng.htm>

## Longitudinal Surveys and Attrition: Issues and Solutions

Charles Victor and Lori Diemert, OTRU

### *The Problem of Attrition*

Longitudinal studies provide researchers with a solid research framework to test the association between multiple exposures and outcomes. However, one of the greatest and most common threats to the validity of the findings of longitudinal studies is attrition. Attrition (i.e., loss to follow-up (LTF)) occurs when subjects cannot be contacted or examined further prior to the planned conclusion of their participation in the study. This leads to missing data on exposures, confounders and outcomes. Attrition threatens the external validity and, in some cases, may threaten the internal validity of longitudinal studies. At the very least, attrition reduces statistical power by decreasing sample size and thus leads to decreased precision of measured estimates. More importantly, attrition leads to selection bias when LTF is dependent on outcomes, potential confounders and/or potential exposures to be examined, which may lead to spurious findings.

In the context of longitudinal telephone surveys, attrition can vary widely. Some telephone surveys report as much as 60% attrition, while Statistics Canada surveys have typically experienced 10–15% LTF. Attrition in telephone surveys may arise from many situations - death or illness, untraceable relocation, change in telephone service, or refusal to continue participation. Strategies aimed at minimizing LTF may be cross-classified based on two dimensions. The first dimension groups strategies based on whether they are active (i.e., the study team attempts to contact the respondent and complete the survey) or passive (i.e., the respondent is given the tools to contact the study team). The second dimension groups these based on whether they are implemented prior to or after identifying a respondent as lost to follow-up.

**Table 1: Strategies to Minimize Attrition by Form and Timing**

		Timing of Strategy Implementation	
		Prior to Loss	Following Recognition of Loss
Strategy Form	Active	<ul style="list-style-type: none"> <li>- Reminder letters</li> <li>- Providing honorarium prior to call</li> <li>- Experienced interviewers</li> <li>- Refusal conversion by experienced interviewer</li> <li>- 12 to 15 callbacks</li> <li>- Home visit</li> <li>- Use of proxies</li> </ul>	<ul style="list-style-type: none"> <li>- Canada 411 Search</li> <li>- Calling alternate numbers</li> <li>- "Google" search</li> <li>- Maintain respondent in call list</li> </ul>
	Passive	<ul style="list-style-type: none"> <li>- 1-800 number</li> <li>- Study website</li> </ul>	<ul style="list-style-type: none"> <li>- Unable-to-contact letter (certified)</li> <li>- Send mail survey</li> <li>- Provide online survey</li> <li>- Advertisement</li> </ul>

One current challenge to active tracking of LTF respondents is privacy laws and ethics. In the past, researchers have used strategies such as tracking respondents through credit bureau reports or through employers; however, except in rare cases, these strategies are deemed unethical and a violation of personal privacy.

### ***Analysis of Longitudinal Data with Attrition***

There is a considerable amount of literature on handling data derived from longitudinal studies in which LTF has occurred. Loss to follow-up data is usually classified as one of the following: 1) missing completely at random (MCAR), where the loss is unrelated to exposures, putative confounders, or outcomes; 2) missing at random (MAR), where the loss is related to an exposure of interest or the putative confounders; or 3) missing not at random (MNAR), where the loss is related to the outcome. In the case of MCAR, there is no bias introduced, and the attrition can be ignored. However, several studies have demonstrated that attrition is most often MNAR, and can lead to serious bias in estimates, even at low levels.

Several strategies to analyze data with missing follow-up values have been proposed. Most of these incorporate some form of imputation, where expected values are imputed for missing values. These expected values may be based on simple means across all respondents, predicted means after adjusting for covariates (i.e., regression imputation, or multiple imputation), or based on weighting, in which the responses from participants with complete information and matching certain characteristics of those lost to follow-up are weighted more highly in the dataset.<sup>1</sup> Finally, many analysts may choose to use only those respondents with complete data in their analysis. Although the imputation procedures add a level of sophistication to the analysis, Kristman *et al.* (1,2) have demonstrated through the use of simulation studies that these methods do not improve on complete subject analysis with respect to the magnitude of the estimate or the standard deviation in the case of MCAR or MAR data. Further, none of the methods employed, including complete subject analysis, can correct for any bias when the data are MNAR.

Some recent studies (3,4) have examined the results of extensive respondent tracking after LTF and have suggested that there may not be differences in early responders compared to late responders. This research serves to remind us that biases due to attrition may not be eliminated by extensive tracking of respondents. This underscores the importance of preventing LTF.

### ***Experience from the Ontario Tobacco Survey***

The Ontario Tobacco Survey (OTS) is a telephone-based, regionally stratified, random sample of Ontario adult (18 years of age and older) smokers and non-smokers. Smokers (any smoking in the past six months) recruited to the baseline survey are followed longitudinally for a minimum

<sup>1</sup> For a pertinent example, see Brown (Appendix A).

of two years at six-month intervals. Respondents receive a \$15 honourarium for each survey in which they are eligible. We described the experience with attrition in the OTS.

### **Attrition Rates**

The highest rate of attrition is observed from baseline to the first six-month follow-up interview; at this time point, we observe a reduction of approximately 14% of the baseline sample. Each subsequent six-month interview resulted in an additional five percent of the baseline sample being LTF. Thus, at the 24-month follow-up interview, the OTS has an attrition rate of approximately 30%.

### **Why do we Lose Respondents?**

In the OTS, LTF respondents were most likely to be untraceable (i.e., we were unable to locate them) or unable to contact (i.e., we were unable to speak with them after the maximum number of call attempts). Small percentages of respondents in each wave were LTF because they refused to continue to participate in the survey (1.5% - 2.5%). On average, less than one percent of respondents were LTF due to 'other' reasons such as being deceased or out of the country.

### **Whom do we Lose?**

In the OTS, a high percentage of LTF respondents were from the youngest age groups: 18-24 year olds and 25-34 year olds. At first follow-up (6 months), a higher proportion of men compared to women were LTF. However, this sex difference was not seen at 12, 18, and 24 months, even though the cumulative proportion of LTF remained higher for men than women. There were no differences in smoking status among LTF at first follow-up; however, at 12- and 18-month follow-ups, occasional smokers were more likely to be LTF than daily smokers or non-smokers. At each six-month follow-up period, there were no education differences among LTF. The implications of LTF in the OTS were discussed.

### **Our Strategies**

The OTS uses a combination of the following strategies to minimize LTF:

1. Reminder letter and honourarium
2. Request for an alternate contact number
3. Toll-free number and website
4. Canada 411
5. Unable-to-contact letters
6. LTFs are not removed from the study until they have missed two consecutive interviews.

Each of these strategies has been successful in capturing respondents that may have been LTF otherwise. The utility of these methods was discussed.

## **References**

Kristman VL, Manno M, Cote P. Methods to account for attrition in longitudinal data: do they work? A simulation study. *European Journal of Epidemiology*. 2005;20(8):657-62.

Kristman V, Manno M, Cote P. Loss to follow-up in cohort studies: how much is too much? *European Journal of Epidemiology*. 2004;19(8):751-60.

Collins RL, Ellickson PL, Hays RD, McCaffrey DF. Effects of incentive size and timing on response rates to a follow-up wave of a longitudinal mailed survey. *Evaluation Review*. 2000;24:347-63.

Clough-Gorr KM, Fink AK, Silliman RA. Challenges with longitudinal survivorship research: attrition and a novel approach of reenrollment in a 6-year follow-up study of older breast cancer survivors. *Journal of Cancer Survivorship*. 2008;2:95-103.

## Methodological Issues in Reporting Cigarette Use and Consumption

Murray Kaiserman, Tobacco Control Directorate, Health Canada

Currently, we are facing a number of methodological issues in conducting surveys, including the following questions:

- Are we using the right metrics to measure smoking behaviours?
- Have we formulated the question correctly, so that it provides meaningful information?
- Are respondents answering truthfully?
- Do respondents understand the question?
- Do we understand what respondents mean with their answers?
- Do we understand the question?

These are not trivial questions as policies, programs and regulations are based upon results from our surveys.

Historically, we have depended upon two major metrics to identify smokers and their levels of smoking – prevalence and amount smoked. What we have always believed is that smokers were telling the truth about their habits. In recent times, however, as smoking has become increasingly socially unacceptable, smokers have been made to feel marginalized and resentful. As a result, we suspect that they may be both reluctant to answer truthfully and to accurately report consumption.

This is leading us to question our metrics. The most used measure has been prevalence. It has been a reliable estimate of smoking behaviour since the first survey in 1964. Or has it? Between 1965 and 1983, overall smoking rates fell from 49.5% to 37.5%<sup>1</sup> and tobacco control advocates were elated. Unfortunately, during the same period, wholesale sales of cigarettes rose from 43 billion units to 63 billion units. Actually, cigarette sales had peaked at 66.5 billion in 1981. This was even more perplexing as self-reported consumption was relatively unchanged and the population grew at a much slower rate than did sales.

The difference between self-reported consumption and cigarette sales was long recognized. Using self-report survey data, if we multiply the number of smokers by the number of cigarettes they report smoking, add all of these numbers and compare with cigarette sales, we see a difference of about 12-13 billion cigarettes in favour of sales. While these cigarettes do not represent all the cigarettes sold during a particular year (some of them are in the distribution pipeline and on retailers' shelves), the discrepancy is huge. This issue has been around for a long time.

Whether in effort to reconcile this difference or not (the rationale for this metric is unnecessary

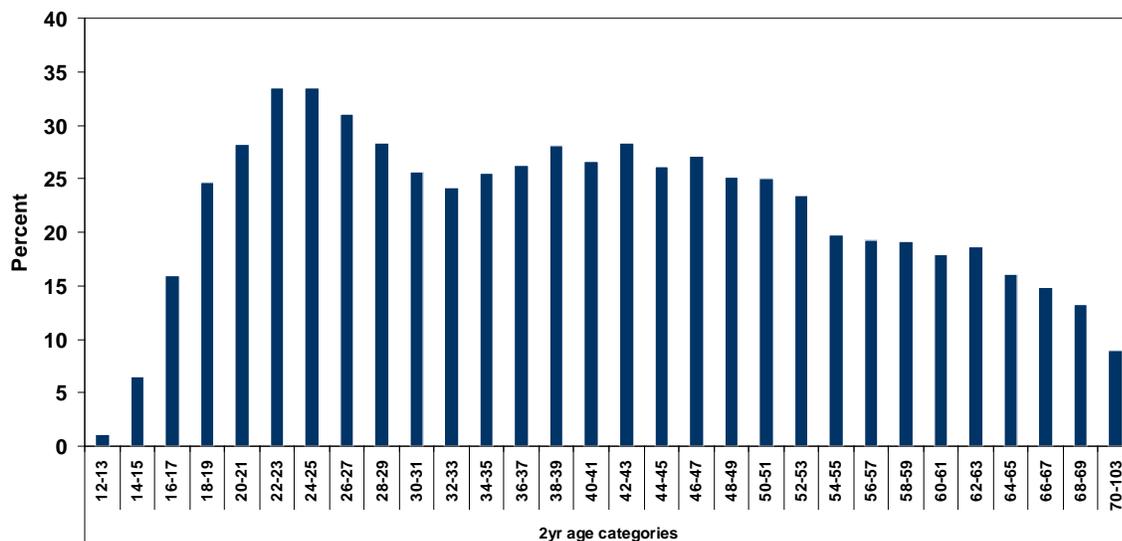
<sup>1</sup> Health and Welfare Canada. Canadians and Smoking: An Update. Ottawa: Minister of Supply and Services Canada, 1991 (Cat. No. H39-214/1991E)

for this presentation), another metric, per-capita consumption, was used. This metric is also problematic, with the question being – what population do you use to calculate per-capita consumption?

This is not a facile question. Wholesale sales data between 1990 and 2007 (2005, 2006 and 2007 are corrected to include estimates of illicit cigarette sales) show that per capita consumption estimates vary both in magnitude and in trend. If one uses the entire population of Canadians aged 15 and older, per-capita annual consumption starts at just under 2,000 cigarettes and trends downwards until 2004 to about 1,500 cigarettes, where it remains flat. If one uses the current smoking population, per-capita consumption remains relatively flat at about 8,100 cigarettes per annum. If one uses only the population of daily smokers, the trend shows an increase from about 9,900 cigarettes to just over 10,000 cigarettes. During this period, per-capita consumption of occasional smokers is not calculated as it is difficult to estimate an average rate of consumption for a heterogeneous group of smokers who do not smoke every day, or even on the same day.<sup>2</sup>

As mentioned earlier, policies, programs and regulations are based upon survey results. But, are we looking at the data properly? If we look at CCHS, 2005, we note that those aged 22-24 have the highest smoking rates at about 33.5% (Fig. 1). There is another peak at about 28% for smokers aged 36-43. Based upon this graph, any additional efforts should target this age group exclusively.

**Figure 1: Prevalence of Current Smoking, by Two-Year Age Group, Age 12+, Canada, 2005**

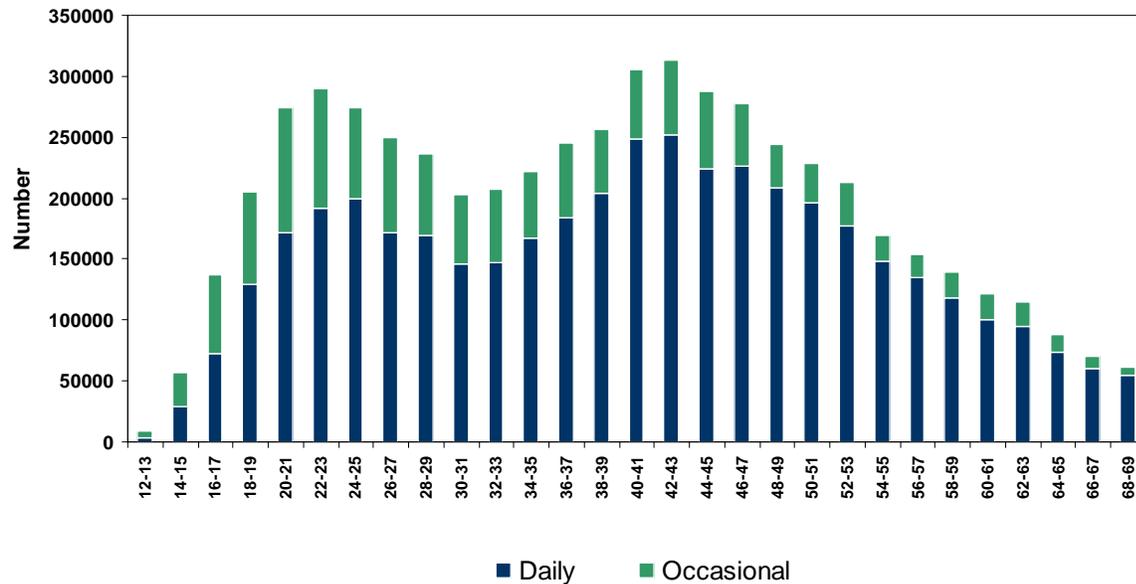


Source: CCHS

<sup>2</sup> Nevertheless, the data do show that, on those days they do smoke, occasional smokers smoked an average of about 3 cigarettes per day.

If we look at the data in a different way, will we reach the same conclusion? Fig. 2 presents the same age groupings, but, instead of prevalence, we use *numbers* of smokers. This chart shows us something very different. Among the smoking population, the age grouping with the highest number of smokers is 40-43. The next largest grouping is young adults aged 20-25. If we use these data, we develop a different strategy – target both older and younger smokers—but in different ways.

**Figure 2: Number of Smokers by Smoking Status, by Two-Year Age Group, Age 12+, Canada, 2005**



Source: CCHS

What these data show is that there is no one metric that can provide a completely accurate picture of what is occurring. In addition, given the current state of smokers and of surveys, validation of responses, or even novel approaches to data collection, might be in order. Finally, the wrong metric, even though we have used it for many years, can lead to the wrong conclusion.

The other questions will be considered using three data sources of illicit cigarette use – CTUMS<sup>3</sup> and two industry studies—the “pack swap study” and the “butt study”.

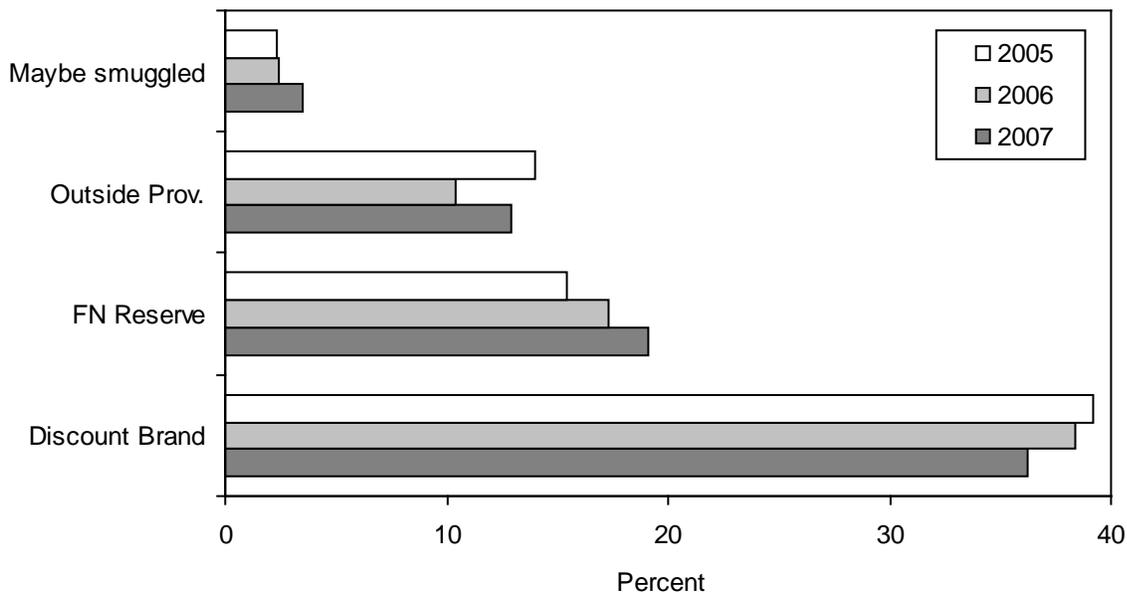
Since 2005, CTUMS has asked smokers about their sources of cigarettes. Over this period, it would seem that First Nations (FN) Reserves as a source has grown to almost 20% of smokers. Looking at Fig. 3, some people might conclude that in 2007, 19% of smokers purchased their cigarettes on FN Reserves and about 35% purchased discount cigarettes. This, of course, neglects to consider the question that was asked:

<sup>3</sup>[http://www.hc-sc.gc.ca/hl-vs/tobac-tabac/research-recherche/stat/ctums-esutc\\_2007-eng.php](http://www.hc-sc.gc.ca/hl-vs/tobac-tabac/research-recherche/stat/ctums-esutc_2007-eng.php)

In the past 6 months, did you buy:

- ... a discount brand?
- ... from a First Nation’s Reserve?
- ... from the internet?
- ... by mail order?
- ... from outside your province?
- ... cigarettes that may have been smuggled?

**Figure3: Self-reported Sources of Cheaper Cigarettes, Current Smokers, Age 15+, Canada 2005-2007**



As written, this question asks smokers for at least one of these activities during the past six months. Responses must be interpreted in this context. What respondents are saying is that during the past six months, they engaged in one, or more, of these activities, at least once, and perhaps regularly. The results cannot be correctly interpreted in any other way.

The industry studies are interesting in that they actually use novel metrics. In the “pack swap study”,<sup>4</sup> the polling firm swapped packs with respondents in order to validate brand use. This is one of the better ways of validating whether smokers are faithfully reporting their brand choices – exchanging packs. Unfortunately, the gains made in the methodological choice are lost in the way the results are reported.

<sup>4</sup> *Illegal Tobacco Sales: A Crisis for Canadians*. National Study for the Canadian Tobacco Manufacturers’ Council. [http://www.imperialtobaccocanada.com/onewebCA/sites/IMP\\_5TUJVZ.nsf/vwPagesWebLive/4611E4A8FE867E17C1256F25004E0898?opendocument&DTC=&SID=](http://www.imperialtobaccocanada.com/onewebCA/sites/IMP_5TUJVZ.nsf/vwPagesWebLive/4611E4A8FE867E17C1256F25004E0898?opendocument&DTC=&SID=)

Using knowledge, attitudes and behaviours of smokers, the study was designed to indicate that use of illicit cigarettes was a problem. As some of the samples from the report show, the study used expressions such as “out of control” and “This is the only national study that measures just how big the problem of illegal tobacco sales is in Canada”. But the study itself has a number of limitations, including that data collection occurred over only one month (was this a “normal” or “unusual” month – see CTUMS); the sample seems to be only smokers (if representative of the population, then the confidence intervals would be much larger as the number of smokers would be about 200-250), yet the report claims to be nationally representative; and the questionnaire is not publicly available.

The “butt study”<sup>5</sup> was also an innovative study; instead of asking youth questions, the polling firm collected their discarded cigarette butts. Unfortunately, this study was misrepresented by the sponsors of the study, who declared that the “Study demonstrates sustained access to illegal cigarettes by 26% of teens in Ontario and 36% in Quebec”. What the study demonstrated was that 26% of the *butts* collected from a small number of high schools in Ontario were illegal. The study did not provide any data with respect to *teens’* smoking. Indeed, the study collected a number of cigarette butts around a small number of high schools during a two-week period. The results are only indicative of what was thrown away by some smokers during those two weeks. The results are not representative of all smokers and of all schools for an entire year.

What we have seen during this presentation is that reports of all survey data must be interpreted with caution. Respondents’ attitudes and understanding affect their answers. Interpretation of results is affected by researchers’ attitudes, understanding and biases. One metric is not enough - we need as many as possible and trends are more important than year-to-year variations. We have to either believe respondents completely, or not - we cannot nor must not pick and choose the answers we like. Even if we believe respondents, it does not hurt to validate. We have to interpret the questions based upon the way they are asked. Finally, when we use novel approaches, we have to acknowledge the limitations of our data – a “pack-swap study” over a two week period is an indication of what some smokers are smoking during that period and not during the rest of the year. Similarly, a “butt study” indicates what was thrown away during the study and may not be representative (only those people who stand outside, may be shared cigarettes, etc).

<sup>5</sup> Canadian Convenience Stores Association. *Youth Contraband Tobacco Study, 2008*  
[www.stopcontrabandtobacco.ca/buttstudy2en.pdf](http://www.stopcontrabandtobacco.ca/buttstudy2en.pdf)

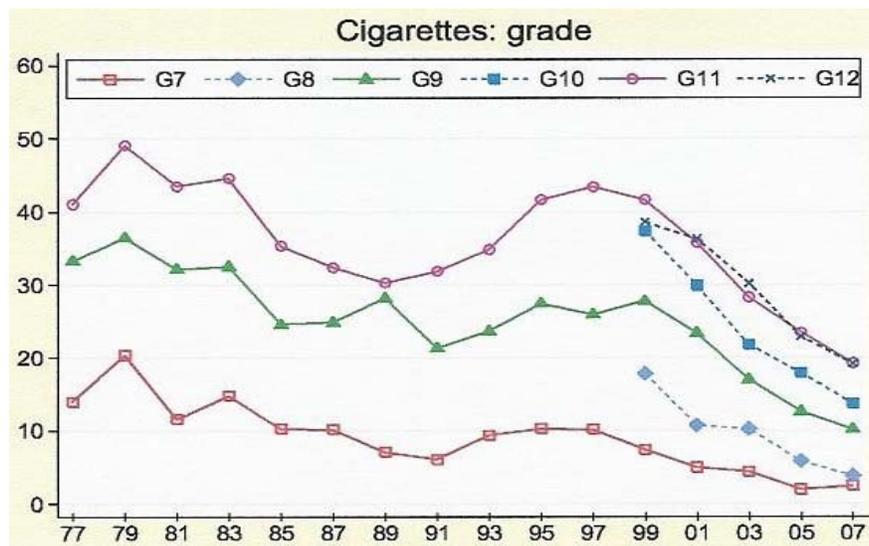
## Comparisons of School-Based and Telephone Surveys of Youth

K. Stephen Brown, Department of Statistics and Actuarial Science, Population Health Research Group, University of Waterloo

### Introduction

The collection of reliable data on youth and its subsequent analysis is very important to tobacco control research and evaluation. Since most cigarette smoking begins by age 19, youth data on 15-19 year-olds provides a window into future adult smoking. While adult smoking rates tend to be fairly constant within five-year age groupings, youth smoking shows considerable variation with single years of age or grade in school. Thus, it is especially important to have enough data that estimates of youth smoking rates may be provided at a fine level of detail. Figure 1 provides one example of the variation in youth smoking by grade and over time. The variation by year can provide clues regarding the impact of policies such as tax cuts and increases. The variation by age indicates that reporting a five-year average rate underestimates the future level of smoking in the population.

Figure 1: Any Smoking in the Past Year by Grade and Year of Survey



Source: Ontario Student Drug Use Survey, 1977-2007

In addition, when youth data are collected at the school level, they can be fed back to schools (e.g. SHAPES, Cameron et al., 2007), identify high-risk schools, and be used as part of natural experiments to evaluate tobacco control initiatives.

### ***Assessing Youth Smoking***

It would be ideal if measures of smoking in adults could be used for youth to present a unified picture of the smoking behaviour of the entire population. However, at younger ages, the focus needs to be on measuring contemplation, initiation, onset, and associated factors, while adult surveys focus primarily on maintenance and cessation. Some adult questions that require estimates of smoking in “the past 30 days” or require counts of the lifetime number of cigarettes have lower test-retest reliability with youth. Further, concepts such as “propensity to smoke” may be more relevant for younger youth than measures based on actual numbers of cigarettes smoked. This poses challenges for a standardized reporting of smoking in youth.

### ***Surveys of Youth Smoking***

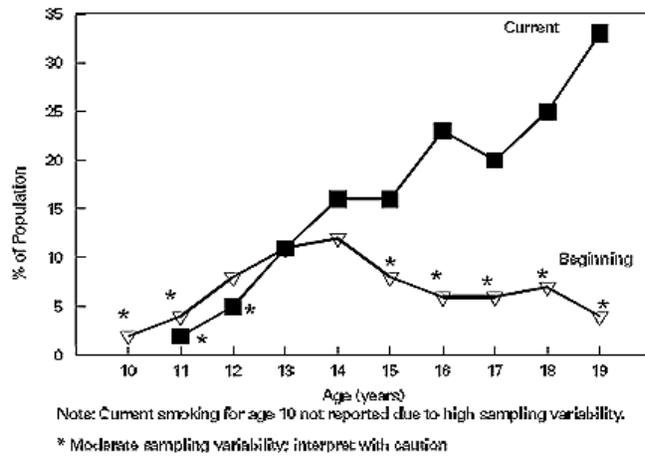
There are a number of national and provincial surveys of smoking that focus on youth under age 20. In particular, the Youth Smoking Survey (YSS) that has been conducted four times since 1994 and is in the field for 2008-2009, has collected data in schools each time. The Canadian Tobacco Use Monitoring Survey (CTUMS), ongoing since 1999, is a telephone survey of Canadians 15 and older that over-samples the 15-19 and 20-24 age groups. In addition there are a number of provincial and regional surveys that collect tobacco use data, often as part of larger drug and alcohol surveys.

The 1994 Youth Smoking Survey (Stephens & Morin, 1996) combined a school-based survey (Grades 5-9; ages 10-14) with a telephone survey for ages 15-19 to get a picture of smoking over this important period of initiation and onset. Figure 2 gives the trend with age for current smokers. At the break between the school component and the telephone component (between ages 14 and 15) there is a curious leveling off of the current smoking rate, followed by a drop in rate at age 17. Whether these are due to methodological issues related to the change in method is not clear.

There are several advantages and disadvantages of school-based and telephone surveys of youth. Telephone surveys provide balanced geographic coverage, more control over survey administration and, theoretically, will cover youth who have left school. However, they have a higher cost per participant, are not feasible with younger students, do not allow for biochemical validation and cannot guarantee the youth’s privacy when responding.

School-based surveys are a relatively efficient way to collect large amounts of data from youth. Data can be collected relatively easily from younger students and they allow for biochemical validation for research grade data. However, they will not survey youth who are no longer in school, or who are absent. Consent must be obtained from school boards, schools, and parents, which lengthens the pre-survey process. Failure to get school board consent can lead to the loss of many schools.

**Figure 2: YSS - 1994 Current Smoking Rates by Age**



Source: Youth Smoking Survey 1994: Technical Report

### **Consent and Response Rates**

Calculation of overall response rates for the two types of survey requires knowledge of the survey design and response rates at each stage in the data collection. The CTUMS survey requires a household to respond and provide a valid roster, with ages of all members. It then requires persons selected from this roster to respond. Thus, the overall response rate calculation requires knowledge of the Household Response Rate and the Person Response Rate. From 1999 to 2007, CTUMS response rates have fallen from the low 80% range to the high 60% range. Interestingly, the Person Response Rate for 15-19 year olds is no different from the average response rate for all age groups.

For the YSS, there have been different sampling procedures over the years; however, all have resulted eventually in a sample of schools being chosen. Then students within those schools are surveyed. This leads to four levels of consent. The school boards/districts that contain the schools must give consent, then the schools, and then the parents (either actively or passively) must consent. Finally students may withdraw on the day of data collection. Replacement boards and schools may be substituted for those not providing consent, but lack of consent for a board or school means the loss of all the relevant students.

The consent process for school-based surveys is often lengthy and requires approval from ethics review committees at all institutions involved in the survey and, increasingly, at the school board itself. School board committees may not meet frequently or their timing may not fit well with the survey timelines, especially without sufficient lead time. School board policies can add conditions, such as active parental consent, that can add costs and delays. At the school level, there is evidence of increasing school fatigue with surveys. In general, with replacement boards

and schools, YSS school consents have reached over 90% for the YSS in each year. Useable questionnaires were received from about 80% of eligible students initially and about 60% in the past two cycles, with a large part of the student non-response coming from active consent schools.

### **Student and Non-student Populations**

CTUMS collects data from both students and non-students through telephone surveys. It might be expected that, for 15-19 year olds especially, the non-student population would be harder to reach, and would smoke more.

**Table 1: CTUMS 2006, 2007 and YSS 2006-07 Student and Non-Student Comparisons – Current Smoking (%)**

Age in Years	CTUMS 2006		YSS 2006-07	CTUMS 2007	
	Student	**Non-Student	Student	Student	**Non-Student
15 - 17	9.7	23.3	10.4	9.4	19.7
18 + *	-	-	16.4	-	-
18 - 19	14.4	36.4	-	17.8	31.1
Total 15-19	11.1	32.8	11.1	12.0	28.1

\* YSS codes only 18 +

\*\* Non- student data based on small sample sizes

Table 1 shows that the current smoker rate for non-students is about twice that of students. With this large difference in smoking rates, it is important that the sample be weighted to reflect actual population sizes. Table 1 also indicates the variation in smoking rate within the 15-19 year age group. Finally, Table 1 indicates close agreement between the student current smoker rates obtained from CTUMS and the YSS.

### **Improving Consent Rates for School-Based Surveys**

Increasingly, it seems that schools are becoming fatigued with surveys. They may not believe that tobacco-use reduction is their mission, or they may feel “used” when they provide data without feedback. Indeed, there are a number of provincial surveys and other initiatives to reduce tobacco use that require school time. Coordinating federal and provincial school-based surveys could be one way to reduce survey burden on boards and schools, and there have been efforts in this regard. The SHAPES program, used in the YSS, provides schools with tailored feedback reports for their school to help schools assess where they stand relative to other schools in the country. With SHAPES, survey modules for physical activity and healthy eating could also be used to

gather data using split-samples within schools to provide schools with a more complete picture of their students.

### **Recommendations**

1. With the rapid change in smoking rates with age in youth, data should be reported by single year of age where possible. At the very least, reporting a single current smoking rate over a broad range of ages should be avoided.
2. Consider weighting CTUMS data to reflect student and non-student numbers.
3. Investigate combining data from CTUMS and YSS to get a more complete picture of 15-19 year olds.
4. Develop a national data base of school-level smoking rates to take advantage of natural experiments.
5. Continue efforts to coordinate federal and provincial school-based surveys.
6. Consider using split-samples within schools to collect data on other behaviours (SHAPES model).
7. Allow sufficient time for school consent process.
8. Develop a coordinated institutional ethics review process across the country.

### **References**

Cameron, R, Manske, S, Brown, K.S., Jolin, M.A., Murnaghan, D., Lovato, C. Integrating public health policy, practice, evaluation, surveillance, and research using local data collection and feedback systems: The example of the School Health Action Planning and Evaluation System. *American Journal of Public Health*, 97(4), 648-654 (2007).

Stephens, T, Morin, M. eds. (Health Canada). *Youth Smoking Survey, 1994: Technical Report*. Ottawa: Minister of Supply and Service Canada, 1996 (Catalogue No. H49-98/1-1994E)

### **Addendum**

Web-based surveys for collecting data from youth and young adults on lifestyle behaviours including tobacco use are an alternative to telephone or school-based pencil-and-paper surveys. Kelli-An Lawrence of Brock University contributed to this session by presenting the results of a study that compared young adult responders and non-responders to an online survey of lifestyle behaviours and use of campus services. (Taylor, L, Lawrence, K-A, 2008)

Web surveys have the advantage that they are less expensive than mail or telephone surveys. They are convenient since the respondent can choose when the survey is completed, and preparation of electronic data files is automatic. Since large numbers of North Americans have

internet access, and teens and young adults may be better reached by web methods than more traditional methods, web-based surveys are a reasonable alternative to more traditional methods for collecting data, especially for younger populations. However, while there are still questions about reliability and non-response bias with on-line surveys, there is increasing evidence that behavioural and psychological measurements commonly used in tobacco research elicit comparable responses whether administered via telephone interviews, paper-and-pencil interviews or web-based interviews.

There are difficulties with open access on-line surveys with determining a sampling frame and, hence, determining response rates. Intercept surveying or list-based approaches do provide a sample frame and for those methods, response rates in the general population are 35-40% (about 13% lower than for mail surveys) and 40-60% for students (about 3% higher than mail surveys). However, in previous studies of smokers' use of the internet and post-secondary students, there were no differences between smokers and non-smokers in either the use of the internet or in responding to a web-based survey (see Taylor and Lawrance for details).

The Taylor and Lawrence study was conducted at 15 Ontario universities. Students entering the campus health centre were given a registration package and, if interested, completed a paper-and-pencil questionnaire. Registrants who accessed the on-line survey within five days were considered responders. Of the 2947 baseline questionnaires disseminated, 2203 (74.5%) were returned. Of those who returned the baseline questionnaire, 1195 (54.5%) accessed the on-line survey. Thus 40.5% of those who took a baseline questionnaire responded. Females were significantly more likely to respond, as were those who had not used cigarettes, marijuana or tobacco alternatives in the past month. There were no differences between responders and non-responders in age, exercise behaviour, alcohol use or the use of on-line communication tools.

In general, web surveys appear to be comparable to mail and telephone surveys in terms of non-response bias, and, in a university aged population, response rates are also comparable. Taylor and Lawrence believe they should be used as the method of choice for health research in the university population. Further research should examine methods such as prompts, contact strategies and incentives that can be used to increase response rates.

### **Reference**

Taylor, L., Lawrance, K-A. A comparison of young adult responders and non-responders to an on-line survey of lifestyle behaviours and use of campus services. Paper presented in the session Addressing Challenges in Tobacco Control Evaluation, *Tobacco Control for the 21<sup>st</sup> Century: Challenges in Research and Evaluation*. Ontario Tobacco Control Unit Conference, Toronto, November 10-12, 2008.

## **Accounting for the Effects of Data Collection Method in Surveys Collecting Data Using Both the Web and the Telephone**

Mary E. Thompson, University of Waterloo

As obtaining a probability sample from a survey population becomes more and more difficult, survey practitioners are turning increasingly to mixed-mode survey methods. Telephone surveying is ideal for many purposes, because the questionnaire is administered by a person trained to elicit information and keep the respondent engaged, and because travel costs are eliminated. Random digit dial methods were developed to overcome the only serious issue with the sampling 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, and to an even greater extent because of access-control technologies such as call-display and automatic screening.

One approach to mitigation of the problem is to use additional frames and data collection modes, to try to increase population coverage and response rates. Self-administered web data collection is particularly attractive because there are no interviewer or data-entry costs. 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. For the most part, web survey frames are lists of email addresses gathered for other purposes of the client. However, many survey firms are developing panels of people willing to respond to surveys on the web for appropriate compensation. The initial recruitment might use telephone or email, and the panels can be described as “rich” frames because personal data of the respondent, including demographics and smoking status, can be collected at recruitment.

The International Tobacco Control (ITC) Netherlands Survey, a survey of adult smokers, oversampling youth, is unlike the other ITC surveys (so far) because most of the participants are responding to the “CASI” (computer-aided self administered) form of the survey on the web. In Wave 1, the target was 1700 CASI participants, and 1800+ were obtained. There was also an RDD component of 400 respondents, included for purposes of assessing the mode effects and facilitating comparison with the ITC surveys in France and Germany, which are conducted entirely by telephone. TNS NIPO, the firm carrying out the fieldwork in the Netherlands, has recruited a panel (essentially a rich frame) of over 200,000 people from the general population for web surveys, and other organizations in the Netherlands have done the same. This is now the way to survey in the Netherlands: our response rate for the telephone component in Wave 1 was 4.6%!

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 the

respondents. For sensitive questions, or difficult questions, or questions for which socially desirable responses exist, it makes a difference whether there is a human interviewer present or not. 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 seen that telephone respondents tend to give more socially desirable responses, and more recently heard responses. They are also more likely to choose the extreme ends of a 5-point Likert scale. These kinds of phenomena are called mode effects, of the technical sort.

Of course, we can only discern these kinds of effects clearly in experiments where respondents are randomized to one mode or the other. I am coming to the belief that ideally, such experiments should be part of major mixed mode studies, and we are attempting one in the ITC Four Country Survey (ITC-4). But of course in the field, this kind of designed experiment can only be approximated. In Wave 7 of ITC-4, we are inviting (by letter) a random sample of longitudinal respondents in each country to respond to the web version, while for the remainder, we will attempt to interview by telephone as usual. However, in our web random sample, there will be some who do not have web access, and some who do not respond for other reasons, and we will attempt to interview them by telephone, but may not succeed. We will not have quite a random sample of web respondents with which to judge the technical mode effects.

It will be helpful in this experiment that we have Wave 6 responses for all sample members, and that many of the questions are the same from wave to wave. We will be able to see differences in the way the responses change, depending on data collection mode, and the differences we see should be mainly technical.

It may well be that in Wave 8, we will give respondents a choice of mode, and in that case, the technical mode effects will be confounded with selection effects. Selection effects will arise because of differences in the characteristics of respondents who will choose one mode or the other. However, the Wave 7 experiment will have given us a good sense of the technical effects.

Another method we have used in some UW Survey Research Centre surveys, in cases where the web frame is a well maintained and there is a comprehensive email list, is to invite all to provide a web response, and follow up non-respondents by telephone. In this scenario, unlike our ITC-4 Wave 7 experiment, the web measuring instrument is thought of as the standard.

The other main kind of mixed mode design is the dual frame design, as we have in ITC Netherlands. Here again, technical mode effects are confounded with selection effects. Although the detailed sampling and recruitment methods for the large web survey panel are proprietary, the frequency tables for the frame show that it is reasonably representative geographically, with a strong representation of younger adults.

To some extent, sampling theory can help with the analysis of dual-frame surveys. In the Netherlands, the telephone frame covers a certain set of residential households, and thus their residents, and we can imagine calculating for each person covered by the frame a probability of inclusion in the telephone sample: the probability of the number being chosen, times the probability of contact being made, times the probability of being selected within the household, times the probability of being available and willing – but roughly, proportional to the number of land lines in the dwelling times the reciprocal of the number of eligible people in the household. The web panel covers a certain set of individuals, and in each age-sex group, for each of those eligible for the survey, there is a (roughly equal) probability of inclusion. The web panel represents a somewhat shadowy portion of the general population, which will overlap with the telephone portion, and together the two frames will account for the union of the two portions. We can partially assess the extent of the overlap by asking the telephone respondents whether they participate or would participate in a web panel, and by asking the web respondents their household composition and how many land-lines they have in their dwelling. Dual frame survey methods (Lohr and Rao, 2000) can then be used to combine the data from both parts of the sample, and guide the construction of weights.

Another useful concept is the “propensity” of response by one mode or the other (Rosenbaum and Rubin, 1983). Theoretically, this is the probability of responding by phone (say), given the fact of being a respondent, as a function of covariates of interest  $X$ , and other covariates  $W$ . It can be shown that, given a particular value of the propensity, the telephone and web parts of the sample are balanced with respect to  $X$ ,  $W$ . We cannot know the propensity, but we can approximate and estimate it using a logistic regression model, regressing an indicator for responding by phone on the covariates  $X$ ,  $W$ . In a sense, the resulting propensity score quantifies selection bias. Controlling for propensity score in comparing the results from the two sample parts allows us to separate technical mode effects from selection bias. Alternatively, comparisons between the two parts of the sample can be facilitated by inverse propensity weighting.

When we come to the point of modelling responses which contain mode effects, the approach will surely depend on the purposes. However, in some cases, it seems reasonable to take the dominant mode to be the standard, and the effect of the other mode to be characterized in terms of parameters in the model. Adding the propensity score to the model may be helpful in making the model parsimonious. For example, we can consider a model of the form

$$\text{logit}(\text{Pr}(Y = 1)) = X\beta + \delta * \text{mode} + \alpha * Z,$$

where  $Z$  is the logit of the estimated propensity to respond by telephone, modelled in terms of the covariates of interest  $X$  and additional variables  $W$ . In this equation,  $\delta$  is a technical effect. With our collaborators in the Netherlands, we will be looking at extensions to ordinal models for items

in the ITC Netherlands survey, where, for example,  $X$  captures age-sex group and education, and  $Z$  depends on these variables, marital status, and the answers to “moderator” questions such as “How often have you been distressed by world events—never, almost never, sometimes, often, very often?” (Telephone respondents in this survey appear much less likely than web respondents to say “Never” or “almost never”.)

Thus statistical methods can be brought to bear on the problems of inference from survey data which are collected using more than one mode. Because mode influences both sampling and measurement, much research remains to be carried out on the most effective implementation of these methods.

### **References**

Lohr, S. L. and Rao, J. N. K. (2000) Inference from dual frame surveys. *Journal of the American Statistical Association* 95, 271-280.

Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.

### **Other useful references (just a sample):**

De Leeuw, E. (2005) To mix or not to mix data collection modes in surveys. *Journal of Official Statistics* 21, 233-255.

Greene, J., Speizer, H. and Wiitala, W. (2008) Telephone and web: mixed mode challenge. *Health Services Research* 43, 230-248.

Schonlau, M., Zapert, K., Payne Simon, L., Sanstad, K., Marcus, S., Adams, J., Spranca, M., Kan, H-J., Turner, R., & Berry, S. (2004). A comparison between a propensity weighted web survey and an identical RDD survey. *Social Science Computer Review*, 22(1), 128-138.

Voogt, R. J. J. and Saris, W. E. (2005) Mixed mode designs. *Journal of Official Statistics* 21, 367-387.

[www.websm.org](http://www.websm.org) (website on web survey methodology)

## Declining Response Rates: The Experience of the International Tobacco Control Policy Project

Christian Boudreau, Department of Statistics & Actuarial Science, University of Waterloo

### *Introduction*

This short paper is comprised of four sections. The first briefly describes the key pitfall of non-response, bias estimation. The second gives an overview of the International Tobacco Control Four Country (ITC-4) Survey. The third is concerned with the cooperation and response rates observed in the first years of the ITC-4 Survey. The last section looks at refusal and non-contact rates, the two components of non-response, to get a better understanding on the potential causes for the declining response rates in the ITC-4 Survey.

### *Effect of Non-response on Estimates*

All surveys aim to estimate, with the greatest possible precision within a budget constraint, quantities (e.g., means and proportions) from a finite population at a given point in time. To this end, complex survey designs, using (for example) stratification and clustering, are employed in combination with (approximately) unbiased estimators (e.g., Horvitz-Thompson estimator). However, survey theory assumes a response rate of 100% when deriving the statistical properties (in particular, unbiasedness and consistency) of these estimators. Practically no survey achieves such a response rate, and consequently estimators may no longer be unbiased. In fact, the lower the response rate, the higher the chances of having a biased estimate and the greater that bias might be.

To illustrate this, consider the case of a simple random sample of  $n$  respondents, where no effort is made to sample non-responders. This is equivalent to sampling from a single stratum (i.e., the one consisting of responders) of a population divided into two strata: responders (consisting of  $N_R$  individuals) and non-responders (consisting of  $N_N$  individuals). In other words, the estimator  $\bar{y}$  is effectively estimating the mean level for the sub-population consisting of the  $N_R$  responders, and not the mean of the overall population of  $N$  individuals. Since  $\bar{y}$  is unbiased for the stratum of responders,  $E(\bar{y}) = \bar{Y}_R$  where  $\bar{Y}_R$  is the mean of the sub-population of  $N_R$  responders. Using the fact that the mean of the overall population and the means of the two sub-populations are related by the following formula,

$$\bar{Y} = \frac{N_R \bar{Y}_R + N_N \bar{Y}_N}{N}, \quad (1)$$

the bias of  $\bar{y}$  is given by

$$\text{bias}(\bar{y}) = \bar{Y}_R - \bar{Y} = \frac{N_N}{N} (\bar{Y}_R - \bar{Y}_N), \quad (2)$$

where  $\bar{Y}_N$  is the mean of the sub-population of  $N_N$  non-responders. Upon closer examination of (2) it is easily seen that  $\bar{y}$  is unbiased if  $\bar{Y}_N = \bar{Y}_R$ ; that is, non-responders have the same mean as responders. This is usually referred to as missing completely at random (MCAR) in the statistical literature (e.g., Little & Rubin 2002). In addition, the bias is independent of the sample size  $n \leq N_R$ , and increasing  $n$  will not yield any bias reduction. Reducing the proportion of non-responders (i.e., reducing  $N_N/N$ ) will, however, reduce bias. An illustration of (2) is given by Levy & Lemeshow (2008), section 13.1.

We conclude section 1.1 by the following two remarks. First,  $\bar{Y}_R$  and  $\bar{Y}_N$  are almost never known in practice. Second, a lower response rate does not necessarily imply greater bias, but will yield a greater proportion  $N_N/N$ , thus increasing a possible difference between  $\bar{Y}_R$  and  $\bar{Y}_N$ .

### ***ITC Four Country Survey***

The International Tobacco Control Four Country (ITC-4) Survey is a prospective longitudinal telephone survey of over 2000 randomly selected adult smokers, in each of the four largest English-speaking countries: Canada, the United States, the United Kingdom and Australia. In this survey, adult smokers are defined as individuals 18 years or older who have smoked more than 100 cigarettes in their lifetimes and who have smoked at least once in the past 30 days. Eligible households were randomly selected using random-digit dialing (RDD). A household was deemed to be eligible if it contained at least one adult smoker as defined above. In households with multiple eligible smokers, the Next Birthday Method was used to select a single respondent. At each yearly wave, cohort members lost to attrition are replaced by newly recruited respondents, to ensure a sample size of at least 2000 per country at each wave. Thus, at each wave, the ITC-4 Survey incorporates both a cohort and a repeated cross-sectional design. The ITC-4 Survey (and the ITC Project in general) uses a quasi-experimental design to measure the psychosocial and behavioural impact of key policies of the Framework Convention on Tobacco Control (FCTC) amongst adult smokers. Further information about the ITC-4 Survey can be found in Thompson et al. (2006).

### ***Cooperation and Response Rates in ITC-4***

The cooperation rate is the proportion of respondents, among those screened to be eligible (i.e., households containing one or more smokers in the case of the ITC-4 Survey), who agreed to complete the survey. There are several formulae for computing cooperation rates. The one used by the ITC-4 Survey (see Table 1) is the AAPOR COOP4.<sup>1</sup> Table 1 gives the cooperation rates for the first six waves of the ITC-4 Survey. Over the first six years of the ITC-4 Survey, cooperation rates have remained fairly stable in the U.K. and Australia and have declined slightly in Canada and the U.S., indicating that, once contact has been established, roughly 75–85% of individuals will agree to complete the survey.

<sup>1</sup> See AAPOR (2000) for more information and alternative formulas.

**Table 1: Cooperation Rates (%) in ITC-4 Survey**

Country	Wave 1 (2002)	Wave 2 (2003)	Wave 3 (2005)	Wave 4 (2005)	Wave 5 (2006)	Wave 6 (2007)
Canada	82.3	84.5	78.8	79.4	75.1	79.9
U.S.	83.2	90.6	74.1	74.9	71.9	73.9
U.K.	78.7	73.8	71.6	78.7	87.3	74.6
Australia	78.8	79.3	79.3	79.9	84.9	79.6

Cooperation rates give a somewhat unrealistic (i.e., too optimistic) picture of the situation, as it is generally impossible to contact or screen a substantial fraction of the target population, and thus determine if these individuals are eligible or not. Examples include individuals who screen their calls, hang-up before the screener, are away from home for the duration of fieldwork or are unreachable for other reasons. These individuals are excluded from the denominator in the computation of the cooperation rate, artificially increasing it. By estimating how many of these unscreened individuals were eligible and including that figure in the denominator, response rates give a more accurate picture of the situation.

Response rates for waves 1–6 of the ITC-4 Survey ranged from 50.0% (Canada – Wave 3) to 12.9% (U.K. – Wave 3), and are given in Table 2 (with the exception that rates for the 2<sup>nd</sup> wave of data collection in Canada and the U.S. could not be obtained from the survey firm that conducted the fieldwork). The response rates of Table 2 are based on a variation of the AAPOR RR4 formula (see AAPOR, 2000). As with the vast majority of surveys, these rates are much lower than the cooperation rates. In addition, after having been fairly high over the first 3 years of the ITC-4 Survey, they have since been declining sharply in Canada, the U.K. and Australia. The decline has been less in the U.S., but the response rate at wave 1 was already of only 25.6%. To gain more insight into these declining response rates, the two components of non-response, refusal and non-contact, are examined in the next section.

**Table 2: Response Rates (%) in ITC-4 Survey**

Country	Wave 1 (2002)	Wave 2 (2003)	Wave 3 (2005)	Wave 4 (2005)	Wave 5 (2006)	Wave 6 (2007)
Canada	49.5	n/a	50.0	29.4	27.3	26.7
U.S.	25.6	n/a	34.9	23.4	20.7	22.3
U.K.	37.8	25.7	41.6	21.4	12.9	15.4
Australia	45.8	36.7	44.2	39.5	45.3	29.1

Note: n/a = not available

### **Refusal and non-contact rates in ITC-4**

There are a few slightly different definitions of refusal rates. We have adopted the one corresponding to the AAPOR REF3 formula; that is, the refusal rate is the proportion of respondents, among those screened to be eligible, who refused to complete the survey or broke-off the interview. The non-contact rate measures the proportion of all cases in the population frame (i.e., adult smokers in the case of the ITC-4) where no responsible member of the household was reached; this definition corresponds to 1 – AAPOR CON2 formula. Refusal and non-contact rates for waves 4–6 of the ITC-4 Survey are given in Table 3. Like cooperation rates (see Table 1), refusal rates have remained fairly stable over the last three years of the ITC-4 Survey; ranging from 1.7% to 7.2%. On the other end, non-contact rates have risen sharply; from an average of 27.4% at wave 4 to 50.1% at wave 6 (i.e., approximately 10% per year). As with many 21<sup>st</sup> Century surveys, the decline in response rates in the ITC-4 Survey is not due to an increase in refusals, but rather to an increase in non-contacts. This phenomenon was noticed by various authors: Steeh et al. (2001), Tuckel & O’Neill (2001), Curtin et al. (2005) and Levy & Lemeshow (2008), section 15.8. Although a more thorough investigation is beyond the scope of this article, we pointed out a few potential causes of this increase in non-contact rates: the proliferation of phone numbers has eroded the efficiency of RDD list, the wide availability of technology like caller-ID now allows individuals to screen their incoming calls,<sup>2</sup> increasing prevalence of cell-phone-only households, the erosion of population receptivity due to incalculable and endless telemarketing campaigns, and the introduction of do-not-call registries in the U.S. and Canada.

**Table 3: Refusal and Non-contact Rates in ITC-4 Survey.**

Country	Refusal Rates (%)			Non-Contact Rates (%)		
	Wave 4 (2005)	Wave 5 (2006)	Wave 6 (2007)	Wave 4 (2005)	Wave 5 (2006)	Wave 6 (2007)
Canada	4.6	7.2	3.8	31.9	50.9	46.1
U.S.	5.4	6.1	2.8	24.4	53.5	47.4
U.K.	4.0	1.7	2.3	29.5	78.5	60.5
Australia	5.5	6.4	4.7	23.8	36.7	46.5

### **Acknowledgements**

The author is grateful to Doctors R. Borland, K.M. Cummings, G.T. Fong, G. Hastings and M.E. Thompson (the PI’s of the ITC Project) for use of the ITC-4 data. The ITC-4 Survey is supported

<sup>2</sup> Tuckel & O’Neill (2001) found that 2/3 of them screen calls “always” or “most of the time”.

by grants from the U.S. National Cancer Institute (Roswell Park TTURC P50 CA111236), the Canadian Institutes for Health Research, the Australian National Health & Medical Research Council, the Australian Commonwealth Department of Health & Ageing, and Cancer Research U.K.

### **References**

AAPOR (2000). *Standard definitions: Final dispositions of case codes and outcome rates for surveys*, American Association for Public Opinion Research (AAPOR), Lenexa, KS.

Curtin, R., Presser, S. & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly* 69, 87–98.

Levy, P.S. & Lemeshow, S. (2008). *Sampling of Populations: Methods and Applications*, 4th ed, Hoboken, NJ.: John Wiley & Sons.

Little, R.J.A. & Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. Hoboken, NJ: John Wiley & Sons.

Steeh, C., Kirgis, N., Cannon, B. & DeWitt, J. (2001). Are they really as bad as they seem? nonresponse rates at the end of the twentieth century. *Journal of Official Statistics* 17, 227–247.

Thompson, M., Fong, G., Hammond, D., Boudreau, C., Driezen, P., Hyland, A., Borland, R., Cummings, K., Hastings, G., Siahpush, M., Mackintosh, A. & Laux, F. (2006). Methods of the International Tobacco Control (ITC) Four Country Survey. *Tobacco Control* 15 (Suppl III), 12–18.

Tuckel, P. & O'Neill, H. (2001). The vanishing respondent in telephone surveys, in *Proceedings of the Annual Meeting of the American Statistical Association*.

## **Appendix B: Participant Questions for the Final Discussion Session**

In preparation for the final discussion, participants at the working session were requested to submit two questions or concerns that were foremost in their minds after 1½ days of presentations and discussion. These questions and issues can be grouped into three themes – measurement and evaluation, response rates, and analysis and dissemination issues. They were used to gauge interest in particular topics and to guide discussion. Not all were addressed. These are listed here for the record as written, with only very minor editing.

### ***Measurement and Evaluation Issues***

- Longitudinal Surveys ~ can they be a source of information for smokers?
  - Has anyone looked at knowledge over time, do people learn? Use the survey to learn? Therefore is this important?
- Do you think bio-markers for smoking (tobacco use) will be used in population surveys in the near future?
- If we use multi-measures (e.g. prevalence of smoking, per-capita consumption) and each shows a different trend, then aren't we increasing the likelihood for someone to report only the partial picture?
- Issues of contraband
  - How to measure it
  - Health Canada's future direction

### ***Response Rate Issues***

- Should we use incentives? Are they making enough of an impact for the cost?
- Would there be any value in increasing the profile of surveys using marketing strategies (ads etc.) in response to the increasing non-response rates/public fatigue?
- Chances of being contacted?
- Mixed mode (as a solution) for attrition
  - Cost
  - Confidentiality requirement
- If mixed modes of recruitment/data collection are recommended for future population-based RDD...
  - What are the optimal modes for recruitment?
  - Is there a maximum number of modes that can be used in a given study?
  - Implications on cost?
  - Implications on response rates and data?
- Comparison of web-based surveys with other modes, e.g. consent issues, attrition rates
- How to determine missing completely at random (MCAR), missing at random (MAR) and missing not at random?
  - What are the solutions for missing data in the above three conditions, especially what can be done for missing not at random?

- Is there any acceptable (minimum) response rate for large surveys, especially for tobacco use surveys?
- Can we change the “perception” of researchers and journals over the “sole” importance of response rates?
- What can be done in terms of survey of cell-phone-only population? Institutionalized population? Homeless population?
- Please clarify what “refusal” conversion is.
- Non-response and its relation to survey bias.
- Recruited panels of smokers, i.e. – “please join our group and we’ll approach you for Public Health stuff and research, so long as you agree to be part of the network”. Should Canadian researchers build one (or several)?

To me, it feels that so far we have mostly been presenting the problems with phone surveys (and web surveys), but so far I have not heard a lot of recommendations for how to deal with those problems and it would be nice to hear several expert opinions on which methods should be used in future and how to properly design surveys to deal with those issues.

### ***Analysis and Dissemination Issues***

- Communicating results to advocates: Is enough being done?
- In longitudinal studies
  - How to treat new data vs. multi-wave data
  - How different are people who stay in - especially regarding their smoking?
- Will bootstrapped weights for CCHS and CTUMS be available for the public use files? When should the bootstrap method be used in survey data analysis?
- Are multiple imputations a solution for missing data?
- If cell-phone only users are different and changing among the target population, how do we weight the data to get population estimates?
- How to deal with reporting web surveys, given they are not representative of the population
  - Given this – challenges inherent with mixed mode surveillance (if you are recruiting from the web)
- Health Canada survey – Re: prevalence
  - Should we use CTUMS vs. CCHS?
  - What are differences & implications?
  - Are they comparable?
- So long as representativeness is maintained, are low response rates that critical? And can we get journals, peer reviewers to be more accepting of low response rates?