USING TOTAL SURVEY ERROR TO STUDY MODE EFFECT AND OTHER APPLICATIONS

Melanie Abeysundera

ABSTRACT

An increasing number of surveys at Statistics Canada have introduced electronic questionnaires as the primary mode of data collection. This change has raised questions regarding the potential differences in response outcomes due to collection mode. Definitions of mode and mode effect are proposed. The relationship between mode and the components of total survey error (TSE) is studied using a proposed extension of the Groves-Lyberg (2010) framework which includes an additional branch for collection. Relationships between the components of TSE and the stochastic mechanisms associated with the sampling process are examined. An example of how the framework was used to evaluate the effect of a change in sampling frame in the General Social Survey will be presented. Lastly, an application of the framework to the testing and the development of the 2016 Census Overcoverage Study will be described.

KEY WORDS: mode effect, total survey error, data collection

RÉSUMÉ

À Statistique Canada, des questionnaires électroniques sont devenus le mode de collecte de données principal dans un nombre croissant d'enquêtes. Ce changement a soulevé des questions concernant les différences potentielles dans les de résultats de participation en raison du mode de collecte. Des définitions de mode et des effets de mode sont proposés. Nous étudions la relation entre le mode et les composants de l'erreur d'enquête totale en proposant d'étendre le cadre de l'erreur d'enquête totale de Groves-Lyberg (2010) qui comprend une branche supplémentaire de collecte. Nous examinons la relation entre les composants de l'erreur d'enquête totale et les mécanismes stochastiques associés au processus d'échantillonnage. Nous présenterons un exemple sur la façon dont le cadre a été utilisé pour évaluer l'effet d'un changement dans la base d'échantillonnage dans l'enquête sociale générale. Finalement, une application du cadre aux tests et au développement de l'étude sur le surdénombrement de 2016 est décrite.

MOTS CLÉS : Effet de mode, erreur d'enquête totale, collecte de données.

1. INTRODUCTION

Over the past few years many Statistics Canada surveys have switched to electronic questionnaire (EQ) as the primary mode of data collection. As a result, questions have been raised with regards to the potential impact this change might have on data quality with respect to total survey error. In 2011, a mode effect working group was formed at Statistics Canada to define and develop guidelines for measuring mode effect. Part of the group's aim was to provide guidance on how to analyse the impact the switch to EQ would have on total survey error. A review of the literature revealed a lack of consensus with regards to the definition of mode effect. Hence, as a first step in understanding how a change in collection mode might affect data quality the terms "mode" and "mode effect" had to be clearly defined (Abeysundera et al. 2012).

A second step, was the realization that mode of data collection is not comprised solely of response related effects (things like selection bias, non-response and measurement error), but also of effects related to the functional aspects of data collection (for example frame development, data processing and follow-up strategies). There are two approaches to total survey error frameworks which predominate in the literature. The first, is to divide total survey error into sampling versus non-sampling components, with a further breakdown into bias and variance components (Smith, 2011; Biemer, 2010). A second approach is to link the components of total survey error to steps in the inferential process associated with a survey (Groves and Lyberg, 2010). Abeysundera and Fox (2012) refined the Groves and Lyberg framework to include an additional set of processes related to data collection. This enabled them to extract components related to mode effect out of the existing framework and to better study the relationship between mode of data collection and the components of total survey error. The proposed framework is versatile and can be used to study other aspects of the collection process.

The remainder of this paper is divided in the following manner. A review of the definition of data collection mode and of mode effect presented in Abeysundera et al. (2012) will be given in Section 2. Section 3 will describe the total survey error framework proposed by Abeysundera and Fox (2012), and review how it can be used to study mode effect. In section 3, some practical applications of the framework beyond the study of mode effect will be discussed. Lastly, some concluding remarks will be given in Section 4.

2. DEFINING MODE EFFECT

There is a lack of consensus with regards to the definition of mode effect within the literature. However, common across all definitions is the concept of a mode-induced difference in the observed response outcomes and an implicit assumption that there is an underlying “true” response (Goves and Lyberg, 2010; Vannieuwenhuize, Looseveldt and Molenbergs, 2010; Schouten, van den Brakel and Klausch, 2011). Furthermore, it is assumed that some portion of the difference between the observed response and this unknown "true" response may be attributed to mode of data collection. It became clear that in order to be able to measure mode effect and determine its impact on survey estimates, we first had to define data collection mode and mode effect. Clearly defining mode effect enables us to gain a better understanding of how a collection mode might affect data quality and to determine how and if this effect can be measured.

Abeysundera et al. (2012) proposed the following definition of **collection mode**. A **collection mode** consists of a combination of three factors:

1. **The collection medium** which is defined as the presence or absence of an interviewer.
2. **The communication medium** which is the instrument through which the questions are communicated to the respondent.
3. **The data recording medium** which is the medium on which the data are recorded (usually paper or computer).

For example, for an electronic questionnaire (EQ) administered via the internet the collection medium is self-enumeration, the communication medium is electronic and the data recording medium is electronic, while for a computer-assisted telephone interview (CATI) the collection medium is interviewer-enumeration, the communication medium is telephone and the data recording medium is electronic.

Based on this definition, a difference in survey outcomes caused by a change in one or more of these factors is then a **mode effect.** To illustrate, consider a survey which collects data on substance abuse. Suppose this survey switches from a CATI collection instrument to an EQ collection instrument. In this case both the collection medium and the communication medium have changed. If the change in the collection medium (presence or absence of interviewer) is responsible for a change in, say average rates of substance abuse reported among our target population, then by the above definition this is a mode effect. It is well known that the presence or absence of an interviewer may influence how a survey unit responds to questions of a sensitive nature, such as questions regarding drug and/or alcohol abuse. In fact, Tourangeau, Rips and Rasinski (2000) found that higher rates of illicit drug use were reported in a self-administered survey than in an interviewer-administered survey.

3. A TOTAL SURVEY ERROR FRAMEWORK

The concept of total survey error was first introduced by Deming (1944), who described thirteen factors which might affect the usefulness of surveys. There has been much discussion in the literature on the application of the total survey error paradigm in the improvement of survey quality (see for example Couper 2011, Smith 2011, Biemer 2010, Groves and Lyberg 2010). Total survey error, as defined by Biemer (2010), is the accumulation of all the errors that may arise from all components of the data collection process. Abeysundera and Fox (2012) expanded the total survey error framework proposed by Groves and Lyberg (2010) to incorporate a branch for data collection. The components in the collection branch proposed by Abeysundera and Fox (2012) were embedded within the other two branches in the Groves and Lyberg framework. By extracting the components related to mode effect out of the existing framework, Abeysundera and Fox (2012) hoped to better study the relationship between mode of data collection and the components of total survey error. Using a detailed schematic, they illustrated how a change in one component in the collection strategy might impact other components in the model and the joint effect they may have on total survey error. Figure 1 shows the modified Groves and Lyberg diagram presented in Abeysundera and Fox (2012).

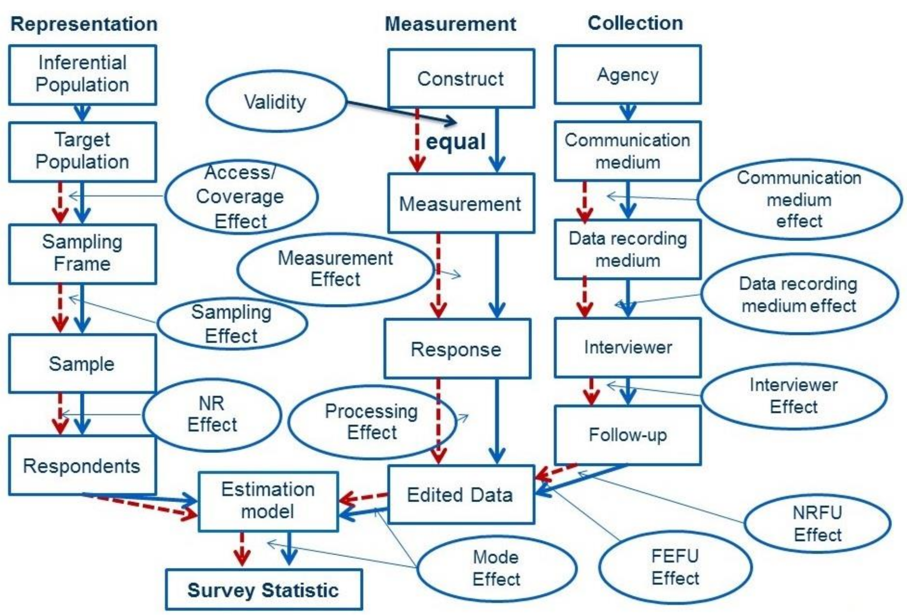

Figure 1. Modified Groves and Lyberg diagram from Abeysundera and Fox (2012).

The inferential processes on the left-hand side in this diagram appear in the original framework by Groves and Lyberg (2010). Groves and Lyberg (2010) divide the survey process into two branches corresponding to the two inferential goals of a survey:

1. To make inferences about a construct or a concept from the measured response of an individual (measurement branch).
2. To make inferences about a target population from a group of respondents (representation branch).

Abeysundera and Fox (2012) added a third branch to encompass components related to mode of data collection. The double arrows correspond to two different modes. These are used to identify components for which a change in mode might induce a mode-dependent effect. Note, that the term effect is used rather than error, where the effect is the difference between two modes at each step in the process. It is important to note that this is not a flow diagram, but rather an arrangement of the interconnected components of the collection strategy employed by the agency. Components with the single arrows are mode-independent and assumed to be fixed across modes. That is, to ensure comparability across modes it is assumed that the agency and the target population are constant across all modes. It is also assumed that the construct being measured is the same regardless of mode. However, if a question is not properly specified a mode-dependent validity effect may occur. For example, suppose that in the first mode a question contained supplementary material which assisted the respondent in answering the question. Now suppose that after a change in mode this supplementary material is either not available or is presented in a different format which leads the respondent to interpret the question differently than he or she did in the original mode. In this example, the concept being measured in the first mode is not necessarily the same as the concept being measured in the new mode, and so there is a difference in response outcomes due to a mode-dependent validity effect in that particular question.

The three main components in the collection branch are **interviewer (collection medium), communication medium** and **data recording medium.** It is these three components which correspond to the three factors which determine mode of data collection as defined in section 2. Based on this definition, a differential effect caused by a change in any one of these three factors is then a mode effect, and a change in any one of these three components will have a rippled effect through the representation, measurement and collection branches of this process. All these components feed into the estimation models and survey statistics so that a mode-dependent difference in the estimates may be observed at the end.

The potential relationship between the various components of this framework can be further examined using Harley-Sielken-Binder-Roberts diagrams. These diagrams illustrate the stochastic mechanisms associated with the sampling process. Figure 2 illustrates how differential access under CATI and EQ might look. The components in this diagram correspond to the representation branch in Figure 1.

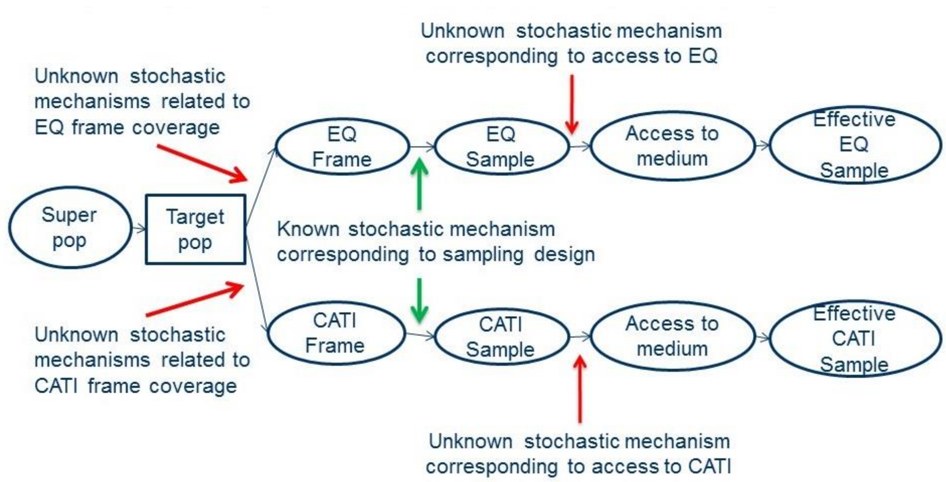

Figure 2. Differential access under computer-assisted telephone interview (CATI) and electronic questionnaire (EQ) modes.

It is assumed that there exists a theoretical superpopulation from which the finite population is generated by some unknown random process. A sample is then drawn from the finite population by a known random process which is the sampling design. However, there are additional stochastic mechanisms related to the mode of data collection within the process. In the example shown in Figure 2, there may be unknown stochastic mechanisms related to frame coverage under the two different modes. If the frame is a list of landline telephone numbers but some members of the target population do not have a landline telephone, than these individuals are not covered by the frame. In addition to this, there may be additional mechanisms corresponding to access to the medium. For example, for EQ surveys the frame is often a list of addresses. A letter with a pass code is then sent out to selected respondents. If some of the sample do not have access to a computer or to an internet connection then those individuals cannot respond. The result is mode-dependent differential coverage in the effective CATI ad EQ samples. However, if the potential sources of this differential access can be identified, than steps can be taken to either minimize their effect or to adjust for it. Further discussion may be found in Abeysundera et al. (2012) and Abeysundera and Fox (2012).

4. APPLICATIONS

The diagrams can be used as a tool to study the relationships between the different components of the collection process and to identify how changes in collection strategies can result in changes in survey estimates through one or many of the components shown. These changes are not necessarily negative. If changes in the collection strategies were introduced to improve the estimates then it is hoped that the observed changes will have a positive impact on survey estimates with regard to data quality. However, changes can also have an unexpected negative impact. A change in collection strategy might cause changes to the coverage of the survey which in turn may result in selection effects if one were to compare the survey estimates under these two strategies.

Miville, Abeysundera and Fox (2014) used the framework to develop a simulation tool to assess the impact of a change in the sampling frame for the General Social Survey (GSS). The simulation study undertaken was a proof of concept that the framework could be used to study the impact of changes to any part of the collection strategy. The GSS has historically used Random Digit Dialing (RDD) to collect cross-sectional data from a random sample of Canadians aged 15 and over living in private households in the 10 provinces. Because the frame was based on landline telephone numbers, cellphone-only households were excluded from the frame. The new GSS sampling frame is based on the common frame for household surveys (MacNabb, St. Pierre and Grenier, 2011). It consists of a list of telephone numbers grouped by address. Addresses are obtained from an address register containing approximately 15 million addresses for private and collective dwellings in Canada. The telephone database is compiled from the 2011 Census of Population, InfoDirect, select mobile phone billing files and tax data. The frame also includes residual telephone numbers which could not be linked to an address. In the new frame some cellular telephone numbers will be included, however landline telephone numbers are still considered the most reliable way to contact people. The telephone numbers for a given address are sorted by reliability, with landline telephone numbers first followed by cellular telephone numbers. Miville, Abeysundera and Fox (2014) created a synthetic population using the 2011 Census of Population and the GSS frame from 2013. Household composition of the synthetic population was compared to that obtained with the Residential Telephone Service Survey (RTSS) 2010, to ensure that the synthetic population was realistic. They then estimated the variable self-rated physical health using data simulated under the old and new sampling frame. Figure 3 shows the GSS sampling strategy under the old (A) and new (B) sampling frames.

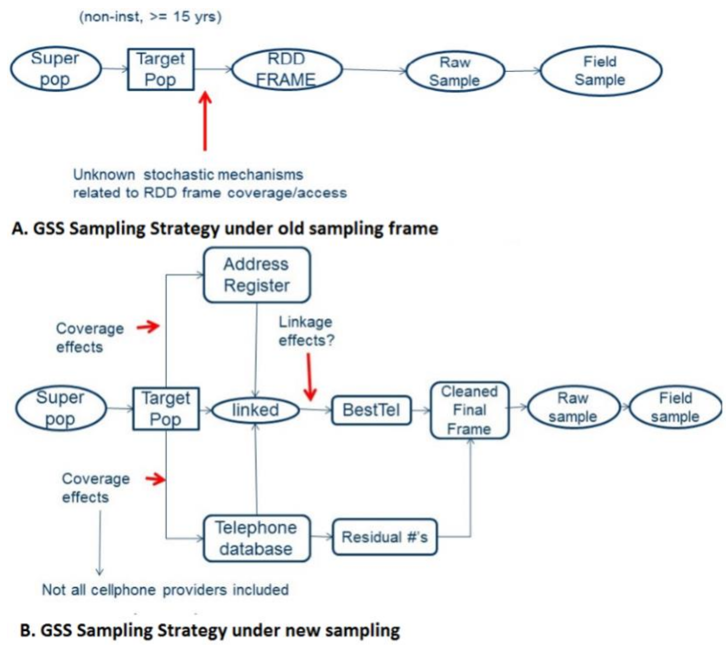

Figure 3. (A) GSS Sampling strategy under old sampling frame, (B) GSS Sampling strategy under new sampling frame.

Using the framework to build their simulation, they were able to replicate what had been observed in the real data. This was a promising result, since it indicated that the framework could be used as a simulation tool to study the impact of changes to the collection strategy of a survey not just with respect to data collection mode, but for other types of changes as well. For example, the simulation tool might also be used to study how a change in sampling design might affect the estimates, or to study how a change in household composition within the target population might affect the estimates.

The framework can also be used as a visual tool to assist in the testing and development of a survey. This next example will illustrate how it can be used to identify the potential effects of changes between the 2011 and 2016 Census Overcoverage Study (COS). The graphs can help to identify where in the process these effects are introduced, and how they might impact related components and final estimates. The potential impact of proposed changes on inter-related components of the process will be tested using 2011 data prior to being implemented in 2016.

The COS is a complete survey which involves many processes including construction of a sampling frame, sampling design, collection, processing, estimation and analysis. Here we will focus on the components related to the creation of the sampling frame. Figure 3 illustrates how the COS frame was built in 2011.

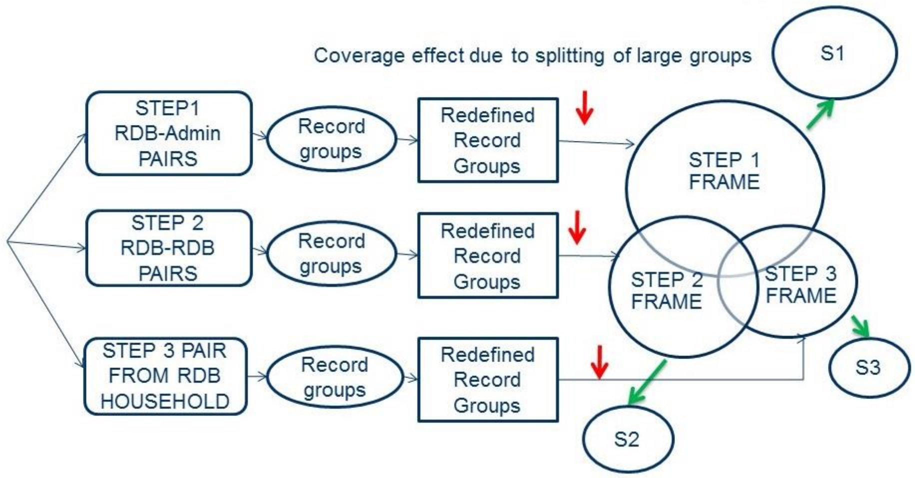

Figure 3. 2011 COS sampling frame

The 2011 frame was built in three separate steps. The first step was a linkage between the census and a large administrative database which covers the entire Canadian population. The administrative frame was built using several data sources including tax data, vital statistics and health care files. The second step was a linkage between those census records which could not be linked to the administrative frame and the entire census. The third was a set of additional linked records obtained using the household identifiers of the pairs obtained in steps 1 and 2. For each step, pairs were placed in interconnected record groups. Because some of these could be quite large, groups were broken into smaller groups, and some links which were considered weak were dropped. This ad hoc splitting of the record groups may have reduced the coverage of the frame as some overcoverage cases may have been dropped from the frame at this stage. Figure 4 illustrates changes that will be introduced to the creation of the COS frame and their impact on different components in the process of frame creation.

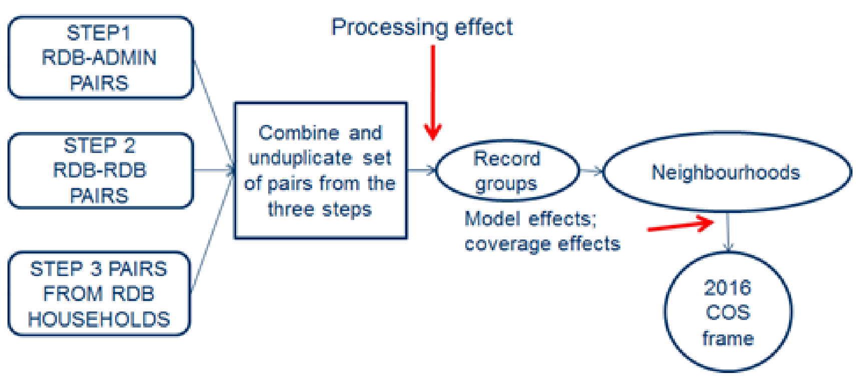

Figure 4. 2016 COS sampling frame

To begin with, the pairs from the three individual steps will be combined and unduplicated and a single set of groups of interconnected record groups will be built. At this stage it is possible that some differences in the record groups have been introduced due to the changes in the procedure used to create the set of record groups. Next, rather than break the groups in an ad hoc manner, a random graph model will be applied to split the groups into neighbourhoods (Dasylva, Abeysundera, Haddou and Lachance, 2015). This new method reduces the size of the individual sampling units while still ensuring all the pairs associated with the original groups remain on the sampling frame. Hence, the new method will improve the coverage of the frame. This is referred to as the coverage effect in Figure 4. The number of neighbourhoods created and their size will depend on the parameters of the random graph model used to do the splitting, and so this will have some effect on the frame produced (the model effect identified in Figure 4). This effect and its relationship to the frame coverage will need to be studied. The graphs enable us to visualize how individual components are related to each other, and how changes to one part of the process might impact other components. We can identify which effects can be studied individually and which will be confounded with other effects. For example, one can see from Figure 4 that changing the way the interconnected record groups are created prior to the creation of neighbourhoods may have an impact on the final frame and that this effect will be confounded with the other effects when comparisons of the two frames are made under the two methods. To isolate the effect of just this component, the process needs to be paused prior to creating neighbourhoods, and the old splitting criteria applied to the record groups obtained with the new method to determine its potential impact independently of the other components. In this manner, the impact of creating record groups from a combined set of pairs can be separated from the impact of the applying the neighbourhood model to split the groups. This will enable us to better understand and explain the results of the testing.

5. CONCLUSION

While the proposed framework was originally developed to study mode effect, it is very versatile and can be a useful tool in identifying where changes to a collection strategy may affect outputs at different stages of the collection process, and how these may affect overall survey estimates. It can also be used to develop test scenarios and simulation studies which can help to identify the effect these changes have on total survey error.

ACKNOWLEDGEMENTS

I would like to thank Karla Fox from Statistics Canada for inviting me to present this paper at the SSC 2015 Annual Meeting. I'd also like to thank Robert-Charles Titus and Christian Olivier Nambeu for reviewing this paper and providing their comments and feedback.

REFERENCES

M. Abeysundera and K. Fox. (2012). "Represent, Measure and Collect: Understanding mode effect within a total survey error framework". *Advisory Committee on Statistical Methods #55.*

M. Abeysundera, J. Claveau, W. Jocelyn, F. Labrecque-Synnott, F. Laflamme, Y. Morin, P. Wright, K. Fox and H. Bérard. (2012). "Mode effects: The medium and the messenger". *Internal Document. Statistics Canada.*

Buelens, Bart and Jan van den Brakel. (2011). "Inference in Surveys with Sequential Mixed-Mode Data Collection". http://www.cbs.nl/NR/rdonlyres/C70ED9D5-6199-4E27-B37D-792A161D614D/0/2011x1021.pdf. *Statistics Netherlands*.

Biemer, Paul P. (2010). "Total Survey Error: Design, implementation, and evaluation". *Public Opinion Quarterly* 74(5): 817-848.

Couper, Mick P. (2011). "The Future of Modes of Data Collection". *Public Opinion Quarterly* 75(5):889-908.

A. Dasylva, M. Abeysundera, B. Akpoué and T. Quadir. (2015). "E-M algorithms for record linkage". *Internal Document Statistics Canada.*

A. Dasylva, M. Abeysundera, M. Haddou and M. Lachance. (2015). "A sampling frame of connected groups for record linkage". *Internal Document. Statistics Canada.*

Deming, W. Edwards. (1944). "On errors in surveys". *American Sociological Review.* 9:359–369.

Groves, Robert and Lars Lyberg. (2010). "Total Survey Error: Past, present, and Future". *Public Opinion Quarterly* 74(5): 849-879.

MacNabb, L, St. Pierre, M. and Grenier, M. (2011). "Development of a common frame for household surveys at Statistics Canada". *Proceedings of Statistics Canada Symposium 2011.* Statistics Canada.

Miville, M.,H., Abeysundera, M. and Fox, K. (2014). "Evaluation of a change of sampling frame in the estimated statistics through simulations technical report". *Internal Document Statistics Canada.*

Smith, Tom W. (2011). "Refining the Total Survey Error Perspective". *International Journal of Public Opinion Research* 23(4): 464-483

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski.(2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.