Public Services and Procurement Canada | Services publics et Approvisionnement Canada

SAGE RESEARCH CORPORATION

THE ADVISORY PANEL ON GOVERNMENT OF CANADA QUANTITATIVE PUBLIC OPINION RESEARCH QUALITY
- Final Report -

Prepared for:
Public Opinion Research Directorate
Public Services and Procurement Canada

Supplier Name: Sage Research Corporation
POR Registration #: POR 058-17
Contract #: EP363-182149/001/CY
Contract Award Date: December 7, 2017
Delivery Date: October 26, 2018
Contract Value: $95,106.45

For more information on this report please contact Public Services and Procurement Canada at:
questions@tpsgc-pwgsc.gc.ca

Ce rapport est aussi disponible en français.

4145 North Service Rd., 2nd Floor, Burlington, ON L7L 6A3 905-577-4040
www.sageresearch.ca

Canada

THE ADVISORY PANEL ON GOVERNMENT OF CANADA QUANTITATIVE PUBLIC OPINION RESEARCH QUALITY

Final Report

Prepared for Public Services and Procurement Canada
Suplier name: Sage Research Corporation
Delivery date: October 26, 2018

This public opinion research report presents the results of a research panel conducted with knowledgeable, leading professionals from the private sector, Statistics Canada and academic institutions, between April 11, 2018 and August 22, 2018.

Cette publication est aussi disponible en français sous le titre : Comité consultatif sur la qualité de la recherche quantitative sur l'opinion publique au gouvernement du Canada

Policy, Planning and Communications Branch
Public Services and Procurement Canada
Portage III Tower A
16A1-11 Laurier Street
Gatineau QC K1A 0S5

Catalogue Number: P103-14/2019E-PDF
International Standard Book Number (ISBN): 978-0-660-28290-9

Related publications (registration number: POR 058-17):
Catalogue Number P103-14/2019F-PDF (Final Report, French)
ISBN: 978-0-660-28291-6 (Final Report, French)

Table of Contents

List of Tables

Executive Summary

Background

The Public Opinion Research Directorate (PORD) is a mandatory common service provider responsible for giving advice on legislation, policies, research methodology and accepted industry practices. Under the Policy on Communications and Federal Identity, PORD also has the responsibility for developing and maintaining Government of Canada Standards. The standards for both telephone and online public opinion surveys were first developed and implemented in 2009 and based on two separate Advisory Panels, one related to telephone and the other related to online research. These standards were later revised in 2013.

Given the ongoing changes in the public opinion research industry, PORD is undertaking a review of its quantitative Standards for the Conduct of Government of Canada Public Opinion Research.

Purpose and Objectives

The project involved convening an Advisory Panel composed of knowledgeable, leading professionals from the private sector, Statistics Canada and academic institutions to provide advice on potential standards and best practices for public opinion survey research (POR) conducted using telephone and/or online quantitative methods.

The Advisory Panel addressed and provided guidance on standards for the following topics:

Sampling
Statistical Treatment of Survey Results
Required Questions in Surveys
Use of Mobile Devices in Online Surveys
Inclusion of Cell Phones and Landline Phones in Telephone Surveys
Telephone Survey Call-back Requirements
Interactive Voice Response (IVR) telephone Surveys
Multi-Mode Surveys
Incentives in Surveys of Children, Young People or Vulnerable Respondents
Privacy and Security of Data
Surveys and Social Media
Accessibility and Literacy

Intended use of the results: For the topics listed above, the intent is to help PORD (a) to revise existing standards and guidelines and, as appropriate, create new standards and guidelines to guide the quality of survey research undertaken on behalf of the Government of Canada, and (b) to equip PORD with expert advice on dealing with evolving research methodologies.

Methodology

The Advisory Panel on Government of Canada Quantitative Public Opinion Research consisted of 10 members drawn from the private sector, academics experienced with market research, and Statistics Canada. Members of the Panel were recruited by PORD, with assistance from Sage Research.

The Advisory Panel's work took place between April 11, 2018 and August 22, 2018. The Advisory Panel process consisted of an initial web conference followed by three online discussion boards. Panel members then reviewed four Working Reports summarizing the results and proposing the guidance to put into the final report of the Advisory Panel.

The Advisory Panel's recommended guidance for quantitative research is expressed as standards and guidelines, together with supporting commentary.

While it was not a mandate of the Advisory Panel to reach consensus, it did so on quite a few aspects of standards and guidelines for quantitative research.

Qualitative research is designed to reveal a rich range of opinions and interpretations rather than to measure what percentage of the target population holds a given opinion. Advisory Panel members gave their personal opinions and experiences on the issues discussed, and were not speaking on behalf of their organization or industry.

Contract Value: $95,106.45, including HST

Political Neutrality Certification

I hereby certify as Senior Officer of Sage Research Corporation that the deliverables fully comply with the Government of Canada political neutrality requirements outlined in the *Communications and Federal Identity Policy* of the Government of Canada andDirective on the Management of Communications. Specifically, the deliverables do not include information on electoral voting intentions, political party preferences, and standings with the electorate or ratings of the performance of a political party or its leaders.

Anita Pollak
President
Sage Research Corporation

The Advisory Panel's Recommendations

Sampling

The Panel provided input on the following topics:

- Definitions of types of samples
- Maximizing representativeness of non-probability surveys
- Online sample information to include in *Proposal Documentation*

There were recommended changes to the following sections of the Standards:

- Section 1 *Proposal Documentation* (1.2.2; 1.2.4 #3; 1.2.6)
- Section 4 *Sampling Procedures* (4.1.1/4.1.2; 4.3.2; 4.4)
- Section 14/15 *Mandatory Survey Report Requirements* (14.7.3/15.7.3)

Definitions of types of samples

Section 4 *Sampling Procedures* includes standards for probability sampling, non-probability sampling and a census but does not give definitions of these types of sampling procedures.

Section 4.1.2 (online)/4.1.1 (telephone) should be expanded (a) to include both definitions and examples of probability and non-probability sampling, and (b) a definition of a census.

Maximizing representativeness of non-probability surveys

The objective is to revise the Standards to emphasize the importance of striving for representativeness in non-probability surveys, and to explain in the proposal how this will be done. The current Standards (sections 4.3.2 *Sampling Procedures* and 1.2.2 *Proposal Documentation*) address this objective to some extent, but the intent is to make the requirement more explicit and detailed.

The majority of the Panel members agreed with including text to emphasize the importance of taking steps to improve the representativeness of non-probability survey results (in both sections 4.3.2 and 1.2.2). However, no consensus was reached on whether this should be a standard or a guideline.

Online sample information to include in *Proposal Documentation*

Section 1.2.4 #3 *Proposal Documentation* lists the required information when an online sample provider is used. This section should be expanded to require separate and more specific proposal information for both probability and non-probability online samples.

There was agreement on the types of information that should be required in the proposal for online probability samples. For online non-probability samples, there were different points of view on some of the specific information disclosure requirements.

Statistical Treatment of Survey Results

The Panel provided input on the following topics:

- Statistical treatment of non-probability survey results
- Statistical treatment of probability survey results
- Statistical treatment of census survey results

There were recommended changes to the following sections of the Standards:

- Section 1 Proposal Documentation (1.2.3)
- Section 4 Sampling Procedures (4.3.3; 4.6.3)
- Section 14/15 Mandatory Survey Report Requirements (14.7.2/15.7.2; 14.7.3/15.7.3)

Statistical treatment of non-probability survey results

The Panel was asked to clarify the use of statistical measures for non-probability surveys in light of recent developments in the application of alternative measures of statistical precision.

Section 4.3.3 of *Sampling Procedures* should include (a) revised wording to further clarify that margins of sampling error do not apply to non-probability survey data; (b) limitations on use of alternative measures of precision (e.g. Bayesian credible intervals) for non-probability surveys; (c) documentation requirements in both the proposal and the survey report when alternative measures of precision are used.

Statistical treatment of probability survey results

The objective was to determine whether to expand on the current requirements for reporting level of precision for probability surveys (section 14.7.2/15.7.2).

There were several proposals for revised wording, but no consensus on the most appropriate wording.

Statistical treatment of census survey results

In section 4.6 in *Sampling Procedures*, there are two statements made about the statistical treatment of census survey results. The Panel was asked to comment on whether the two statements were consistent.

It was agreed that part of section 4.6.3 should be deleted as it is inconsistent with the standard to not use inferential statistical tests in a census survey.

The Panel was also asked to consider whether or not a survey ceases being a census if the response rate falls below a certain level.

There was agreement that a census with a response rate of less than 100% is still a census, albeit perhaps better described as an attempted census. Margin of sampling error does not apply but other sources of survey error can still be present, such as non-response bias.

Required Questions in Surveys

The Panel provided input on the following topics and recommended changes to Section 2 *Questionnaire Design* (2.1.2; 2.1.3):

- Introduction wording preceding required demographic questions at the end of a questionnaire
- Use of *prefer not to answer* in the required questions for online surveys
- Efficiency of reading a large number of response options for required questions in telephone surveys
- Required demographic questions:
  - Gender
  - Language
  - Age
  - Education
  - Household Income
  - Addition of household phone status for telephone surveys

Introduction Wording preceding required demographic questions

There is a general requirement (section 2.1.2) to inform respondents at the beginning of a survey of the confidentiality of their questionnaire responses, but the current standards do not state any specific wording for how to preface the block of demographic questions located near the end of the questionnaire.

Add a requirement for an introduction to the block of demographic questions that addresses confidentiality. There was agreement on most of the wording for this introduction, but no consensus was reached on use of the terms “confidential” and/or “anonymous” in this introduction.

Use of *prefer not to answer* in the required questions for online surveys

Prefer not to answer (and its related forms) should be removed as a required listed response category in online surveys, but retained as an optional response category.

Efficiency of reading a large number of response options for required questions in telephone surveys

Some of the required demographic questions have a relatively long list of response options. This includes the questions on age, education and household income.

Revise section 2.1.3 to allow for modifying the wording of a demographic question to allow the interviewer to instruct a respondent to stop at the category that applies to them.

Required demographic questions: Gender

In the currently mandated question, gender information is collected very differently in telephone versus online surveys. The telephone version does not actually ask a question about gender, but rather relies on interviewer observation. The online version asks a question and so is based on respondent self-classification rather than interviewer classification. Neither the telephone nor online survey versions offer an “other” answer option.

Revise the mandated question for all POR surveys (telephone and online) (a) to include an "other" answer option, and (b) require the gender question to be read to respondents in telephone surveys.

Required demographic questions: Language

There are currently two mandated questions for language – mother tongue and language spoken most often at home, with discretion for the researcher to use one or both questions depending on the survey objectives.

Revise this to require only language spoken most often at home.

Required demographic questions: Age

Revise two of the current age categories (35-49 and 50-54) to 35-44 and 45-54 to get a more even distribution of the age categories.

There was discussion but no consensus on whether the current age category 18-34 should be split into two categories (18-34 and 25-34) or left as is.

Required demographic questions: Education

In order to better align the response options with the school systems in both Quebec and the rest of Canada, combine "Grade 8 or less" and "Some high school" into a single category, "less than high school diploma or equivalent."

Required demographic questions: Household income

Revise the question wording to specify a time frame of "last year" for household income.

Required demographic questions: Addition of household phone status for telephone surveys

This information can sometimes be useful in quota controls or weighting. Add a requirement to section 2.1.3 (a) to include a question on household phone status in all telephone surveys.

Use of Mobile Devices in Online Surveys

The current standards do not address the possibility and implications of an online survey being completed on a mobile device. The Panel was asked to provide input on revising the Standards to address the following areas:

- Proposal documentation relating to use of mobile devices in online surveys
- Mobile-friendly online surveys and questionnaire design
- Proposed revisions related to *Pre-testing* in the Online Standards
- Possible revisions to Online Standards related to *Data Collection* and *Quality Controls* related to the possibility of mode effects by device type/screen size
- Covering respondent costs for use of mobile devices

There were recommended changes to the following sections of the Online Standards:

- Section 1 Proposal Documentation (1.2.2; 1.2.5; 1.2.7)
- Section 2 Questionnaire Design (2.1)
- Section 3 Pre-testing (3.1)

- Section 7 Data Collection (7.7)
- Section 14 Mandatory Survey Report Requirements (14.6)

Proposal documentation relating to use of mobile devices in online surveys

The default expectation should be that an online POR survey sample will include respondents using either a computer or a mobile device for the survey and that surveys have a mobile-friendly version of the questionnaire.

Additions to sections 1.2.2 and 1.2.5 in *Proposal Documentation* are recommended to make these expectations explicit.

Mobile-friendly online surveys and questionnaire design

There were three potential revisions/additions to Section 2 *Questionnaire Design* considered with respect to questionnaire design in online surveys where mobile devices may be used.

Should there be a standard encouraging use of a common question design/layout across devices?

The consensus was that a standard is not appropriate given that research on what design approach is best for question design/layout across devices is inconclusive, and optimal design approach can vary across surveys and for different questions within a survey. However, the addition of a guideline highlighting the options available to researchers could added in Section 2 *Questionnaire Design*.

Should there be a different survey duration standard for mobile-friendly surveys?

*T*he standard for online questionnaire duration is 20 minutes, but an average duration of 15 minutes or less is "strongly encouraged." The Panel considered whether the standard for survey duration should be left as is, or revised to specify a shorter duration for mobile-friendly surveys.

No change was recommended to the existing standard on survey duration.

Should there be guidelines on features of a mobile-friendly questionnaire?

Most agreed that a list of examples be added to Section 2 *Questionnaire Design* as a useful reminder to researchers of the elements that make a questionnaire more mobile-friendly.

Proposed revisions related to *Pre-testing* in the Online Standards

The current standard specifies the total number of pre-test completions, but does not break this down by device type.

The Panel considered potential revisions/additions to Section 3 *Pre-testing* with respect to online surveys where both mobile devices and computers may be used.

Should pre-testing standards specific to device type be added?

Section 3 *Pre-testing* should be revised to include a requirement for pre-testing on both computers and mobile devices when a survey can be completed on both types of devices. Several alternative options for how to word the requirement were proposed.

Should standards on the number of pre-test interviews by device type be added?

The consensus was that section 3.1.5, which requires a minimum of 10 pre-test interviews in each language, should be left as is with the understanding that the pre-test would include a sample of different devices.

Possible revisions to Online Standards related to *Data Collection* and *Quality Controls* related to the possibility of mode effects by device type/screen size

In a survey that allows completion on both mobile devices and computers, there is the potential for a "mode" effect. That is, the different designs/layouts for a given question could cause different response distributions.

The Panel considered whether or not there should be any requirement to collect information on device type, and any requirement to conduct an analysis for mode effects by device type.

A standard should be added to Section 7 *Data Collection* requiring collection of data on the type of device used by respondents to complete a survey.

The Panel did not support adding a standard requiring an analysis of mode effects for each survey (section 14.6 *Quality controls*). The view was that not enough is known about device type mode effects at this time to specify analytic requirements for individual surveys and "research on research" needs to be done using the aggregated device data collected across surveys to determine what if any standard would be appropriate for analysis of potential mode effects.

Covering respondent costs for use of mobile devices

Users of mobile devices may incur costs to participate in a research survey. The current standards do not have any requirements as to how such costs should be handled.

The consensus was that there should not be a standard about covering respondent costs associated with using a mobile device: (a) respondents always have a choice whether or not to participate in a GC POR online or telephone survey; (b) the current standards require certain information be given about the survey (e.g. length), so respondents are able to make an informed choice about whether or not to participate; (c) unless compensation is set at an arbitrary fixed amount for all mobile users, the logistics of determining the amount to compensate each respondent and documenting this for billing purposes would be very complex and difficult, if not impossible.

Inclusion of Cell Phones and Landline Phones in Telephone Surveys

An important issue in sampling for telephone surveys is the inclusion of cell phone users and landline users. This can affect coverage of the survey population, the sampling frame(s) used for the survey, and possibly weighting. A telephone probability sample of the general Canadian adult population must include a cell phone sample. The Panel was asked to consider revisions to the Telephone Standards related to *Proposal Documentation* and *Sampling Procedures*:

There were recommended changes to the following sections of the Standards:

- Section 1 *Proposal Documentation* (1.2.3 #1; 1.2.4 #7)
- Section 4 *Sampling Procedures* (4.2.3c)
- Section 15 *Mandatory Survey Report Requirements* (15.5)

Proposal documentation relating to inclusion of cell and landline phones in telephone surveys

There were three potential revisions/additions to Section 1 *Proposal Documentation* in the Telephone Standards considered by the Panel.

Response rate/participation rate

Consider revising the text in section 1.2.3 #1 to require stating an estimated response/participation rate for both cell phones and landline in surveys where both device types can be used.

Description of data collection

The Panel was asked to consider whether there should be any revisions to section 1.2.4 #7 in *Proposal Documentation*, which states that a rationale must be given when the sample includes interviews on cell phones. The current language overly downplays the importance of including cell phone users in the sample.

The wording of section 1.2.4 #7 should be revised to acknowledge the importance of cell phone samples in telephone surveys. There were several alternative proposals on the approach to take.

Sampling procedures relating to inclusion of cell and landline phones in telephone surveys

The current standard in *Sampling Procedures* section 4.2.3c addresses disclosure of coverage issues in probability samples, and gives as an example a sample of cell phone only households.

Add landline-only samples as another example in section 4.2.3c given the growing number of cell phone only households, a landline-only sample could have substantial coverage error.

Telephone Survey Call-back Requirements

The Telephone Standards for call-backs in Section 7 *Data Collection* (7.2) require a minimum of eight call-backs to be made before a telephone number is retired. Some concern has been expressed that eight call-backs is too many, and might be perceived as harassment. The current standard also (a) does not provide a definition of what constitutes a call-back, and (b) does not differentiate between call-backs to landlines and cell phones. The Panel was asked to consider what should be the standard for number of call-backs, including whether there should be a different standard for respondents reached on a cell phone.

There were two main recommendations for revisions to section 7.2:

1. Change the terminology from “call-backs” to “call attempts” on the grounds the meaning is more straightforward. Note that “call attempts” equals one plus the number of “call-backs.”
2. A minimum of eight call-backs (nine call attempts) is excessive. The majority Panelists recommended the Standard be revised to require six call attempts, meaning the initial call and five call-backs.

The Panel opted to apply the same call-back requirement to both cell phones and home phones.

Interactive Voice Response (IVR) Telephone Surveys

The Panel was asked to consider revisions to the Telephone Standards in section 5.3 *Use of Interactive Voice Response* in the following areas:

- Revisions related to the use of IVR surveys for GC POR surveys
- Requirements for an IVR survey introduction

The Panel was also asked to provide input on standards related to:

- IVR survey duration

- Whether there should there be a different call-back standard for IVR surveys

There were recommended changes to the following sections of the Standards:

- Section 2 Questionnaire Design (2.1.1)
- Section 5 Use of Interactive Voice Response (5.3.1; 5.3.2)
- Section 7 Data Collection (7.2)

Use of Interactive Voice Response for GC POR surveys

Section 5.3.1 discourages, but does not forbid, use of IVR surveys for POR. It also suggests circumstances when IVR may be an appropriate methodology. The Panel considered whether there should be any changes to this sub-section on the *Use of Interactive Voice Response*.

The majority suggested adding more examples of situations when IVR as a data collection method may be acceptable while maintaining the principle that IVR is not a preferred method for GC POR surveys.

IVR Survey Introduction

Section 5.3.2 states that the information disclosure requirements for IVR surveys are the same as for interviewer-administered surveys, and similarly requires that the information be provided in the survey introduction. Because IVR surveys are typically shorter than interviewer administered surveys, the Panel was asked to comment on (a) whether the required elements for telephone survey introductions should be revised or shortened for IVR surveys, and (b) the possibility of moving some of the information disclosures to the end of the survey.

Most Panelists said the required information in the survey introduction should be the same for IVR surveys as for other surveys. There was no consensus on where in the questionnaire the various required disclosures should be made. However if it is decided that certain types of information can be disclosed later in a survey, that option should be available to all surveys, and not just to IVR surveys.

IVR survey duration

The standard for survey duration states surveys must be completed in 20 minutes, and strongly encourages a duration of 15 minutes or less.

A guideline should be added to *Questionnaire Design* section 2.1.1 to encourage an IVR survey duration of five to seven minutes or less.

Should there be a different call-back standard for IVR surveys?

The call-back requirements in section 7.2. *Call-Backs* do not make any distinction between interviewer-administered surveys and IVR surveys. The Panel considered whether there should be any changes to this section specific to IVR surveys.

There were several alternative proposals, ranging from a recommendation to exempt IVR surveys from the call-back requirements for interviewer-administered surveys, to requiring IVR surveys to have the same call-back requirements as interviewer-administered surveys.

Multi-Mode Surveys

The current standards already address multi-mode surveys to some extent. The Panel was asked to provide input on possible revisions to the Standards in the following areas related to multi-mode surveys:

- Proposal documentation
- Sampling procedures
- Questionnaire design
- Pre-testing
- Reporting outcome rates
- Mandatory survey report requirements

There were recommended changes to the following sections of the Standards:

- Section 1 Proposal Documentation (1.2.4 #7; 1.2.7 #1)
- Section 2 Questionnaire Design (2.1)
- Section 3 Pre-testing (3.1)
- Section 4 Sampling Procedures (4.5)
- Section 8 Outcome Rates
- Section 14/15 Mandatory Survey Report Requirements (14.5.2/15.5.2; 14.6.3/15.6.3)

Proposal documentation for multi-mode surveys

The primary concern associated with multi-mode surveys is the potential for mode bias. The Panel considered whether and how the proposal documentation requirements need to be elaborated to make it more clear in the proposal that the issue of potential mode bias is recognized and that steps will be taken to address this.

Additions were recommended to:

- Section 1.2.4 #7 (*Description of Data Collection*): Add a requirement to give a rationale for the modes that will be used, and describe the steps that will be taken to reduce the likelihood of mode biases and to facilitate detection of any mode biases.
- Section 1.2.7 #1 (*Data Analysis*): State whether the plan is to combine data across modes or to report the results separately by mode.

Sampling procedures for multi-mode surveys

Revise section 4.5 in *Sampling Procedures* to emphasize the value of using similar modes of data collection to minimize the risk of mode biases.

Questionnaire design for multi-mode surveys

There is no current standard for *Questionnaire Design* specific to multi-mode surveys.

Revise section 2.1 in *Questionnaire Design* (a) to encourage comparability across modes in question wording and presentation of response options, and (b) to highlight the value that including benchmark questions can have for enabling detection of mode biases.

Pre-testing for multi-mode surveys

The current Section 3 *Pre-testing* does not make any specific references to separate pre-tests by mode in a multi-mode survey. The Panel was asked to consider whether there should be a requirement for a minimum number of pre-test interviews in English and French for each mode in a multi-mode survey.

There were several different points of view on this matter, and no consensus was reached.

Outcome rates for multi-mode surveys

Currently in Section 8 *Outcome Rates* there is no standard for how to calculate outcome rates for a multi-mode survey.

A standard should be added outlining the general principles for calculating and reporting on outcome rates for multi-mode surveys. Research designs which do not allow calculation of either of the mandatory outcome rates (response rate or participation rate) should not be permitted for GC POR surveys.

Mandatory survey report requirements for multi-mode surveys

The standard for reporting on data collection in sections 14.5.2/15.5.2 of *Mandatory Survey Report Requirements* should be updated using the updated language in sub-section 1.2.4 #7 in *Proposal Documentation*.

Section 14.6.3/15.6.3 *Quality Controls* should be revised (a) to ensure decisions made about combining or not combining data across modes are clear, and (b) to require descriptions of any adjustments made to the data to mitigate mode effects.

Incentives in Surveys of Children, Young People or Vulnerable Respondents

Section 6 *Data Collection from Children, Young People or Vulnerable Respondents* does not make any reference to whether or how incentives are used for this survey population. Section 7 *Data Collection* (7.5 [telephone]/7.6 [online]) that deals with incentives/honoraria also does not refer to this population.

Guidance should be added to section 7.5/7.6 to address such matters as who will receive the incentive and getting parental consent.

Privacy and Security of Data

The Panel was asked to provide input on possible revisions/additions to the Standards in the following areas:

- Passive data collection in online surveys
- Respondent photographs and recordings
- Telephone surveys – Sensitivity to setting
- Data breaches
- Cloud storage

There were recommended changes to the following sections of the Standards:

- Section 2 Questionnaire Design
- Section 7 Data Collection (7.2)
- Section 13/14 Data Security

Passive data collection in online surveys

Online and mobile methodologies create possibilities for collecting various types of personal data "passively" that is, without direct interaction with respondents. The issue considered was what passive data collection is allowed and under what circumstances is it allowed in the context of surveys? The Panel was asked to consider if the current standards are sufficient to address these questions associated with passive data collection in surveys.

The Panel endorsed a revision to the standards in *Data Collection* section 7.2 (a) to explicitly define "passive data collection" and provide examples of "personal information", and (b) to note exceptions where the passive data collection is legally permissible.

Photographs and recordings

The Online and Telephone Survey Standards do not currently have any standards pertaining specifically to respondent photographs, videos or audio recordings.

The Panel endorsed the addition of standards in Section 5 *Retaining Public Confidence* to clarify (a) that photographs and recordings are considered to be personal data and need to be treated as such, and (b) the responsibility of researchers when the survey involves asking respondents to generate photographs and/or recordings.

Telephone surveys – Sensitivity to setting

The current Telephone Standards section 5.2.1 *Avoidance of Harassment*, has a standard focused on sensitivity of the survey subject matter, but it does not directly address issues potentially caused by the setting of the interview. Because respondents are increasingly likely to answer calls using a mobile phone, there can be issues with them using the phone in problematic settings (e.g. driving, walking in a public space). On both mobile phones and fixed-location phones, they may be in a setting where they can be overheard.

Most Panelists supported adding a guideline to determine if a telephone survey respondent is in a location where they can take the call, for both cell and landline users (Section 2, *Questionnaire Design*).

Data breaches

The current standards in Section 13/14 *Data Security* require taking steps to protect against data breaches (the loss of or unauthorized access to/disclosure of personal or organization information). The relevant sections are: 13.2 (online)/14.2 (Telephone) *Protection of data servers*; 13.3/14.3 *Temporary storage of data on servers*; (c) 13.6/14.5 *In the event of any data breach*.

The Panel was asked to identify any revisions or additions to the standards, and/or any guidelines that should be included.

The existing standards pertaining to privacy and security of data, including data breaches, are appropriate.

There were two main areas identified for additional standards or guidelines:

- Encryption of data on portable devices: Add a standard requiring encryption of survey data files on portable devices that contain personally identifiable information (PII), and add a guideline to encrypt when there is not PII.
- Data retention limits: There is no standard requiring destruction of PII after a period of time. The Standards should include rules about retention limits for PII, but there were different points of view on (a) whether this should be a standard or a guideline, and (b) how the time frame should be expressed.

Cloud Storage

The current standards in Section 13/14, *Data Security* require that survey data be stored in Canada.

This is a complex area: it requires expertise in the legal and regulatory framework affecting data access and use not only in Canada but in other countries as well where servers might be located, and it requires an understanding of GC policies in this area. The Panel did not consider itself to be experts in these areas. For the most part there were no suggested changes to the current standards. However, one suggestion was for the GC to have a pre-approved list of countries that satisfy the conditions set out in in the current standards and that are acceptable for cloud storage of GC POR data.

Surveys and Social Media

The Panel considered whether there are any additional standards required for surveys that use a social media venue as either a sample source or to administer a survey.

The current standards, together with the various changes recommended elsewhere by the Panel, are sufficient to ensure that any such surveys meet the quality requirements for GC POR surveys. Therefore no additional standards are needed for surveys that use a social media venue for either sampling or survey administration.

Accessibility and Literacy

The Online and Telephone Standards do not contain any standards or guidelines pertaining to accessibility.

The Panel considered whether a statement should be added to the Standards about the importance of accessibility, and what if any specific guidelines might be provided for online and telephone surveys. Note that according to PORD, the Treasury Board Secretariat (TBS) is working on a proposed policy for accessibility standards specific to all devices used to access online surveys. The results of this development work will probably be available in a year or so. When the TBS policy is finalized, it will take precedence.

The majority of Panelists supported adding a general guideline encouraging accessibility including providing guidelines on some examples of steps that could be taken to improve accessibility in online or telephone surveys.

Introduction

Background

The Public Opinion Research Directorate (PORD) is a mandatory common service provider responsible for giving advice on legislation, policies, research methodologies and accepted industry practices. Under the Policy on Communications and Federal Identity, PORD also has the responsibility for developing and maintaining Government of Canada Standards. The standards for both telephone and online public opinion surveys were first developed and implemented in 2009 based on two separate Advisory Panels, one related to telephone and one related to online research. These standards were later revised in 2013.

Given the ongoing changes in the public opinion research industry, PORD is undertaking a review of its quantitative Standards for the Conduct of Government of Canada Public Opinion Research.

Purpose and Objectives

The project involved convening an Advisory Panel composed of knowledgeable, leading professionals from the private sector, Statistics Canada and academic institutions to provide advice on potential standards and best practices for public opinion survey research (POR) conducted using telephone and/or online quantitative methods.

The Advisory Panel addressed and provided guidance on standards for the following topics:

Sampling
Statistical Treatment of Survey Results
Required Questions in Surveys
Use of Mobile Devices in Online Surveys
Inclusion of Cell Phones and Landline Phones in Telephone Surveys
Telephone Survey Call-back Requirements
Interactive Voice Response (IVR) telephone Surveys
Multi-Mode Surveys
Incentives in Surveys of Children, Young People or Vulnerable Respondents
Privacy and Security of Data
Surveys and Social Media
Accessibility and Literacy

Intended use of the Results

For the topics listed above, the intent is to help PORD (a) to revise existing standards and guidelines and, as appropriate, create new standards and guidelines to guide the quality of survey research

undertaken on behalf of the Government of Canada, and (b) to equip PORD with expert advice on dealing with evolving research methodologies.

Manner in which research is prescribed by legislative, policy, evaluation or litigation requirement: PORD provides coordination and advisory services for public opinion research (POR). These services are mandatory for POR contracted by institutions listed in schedules 1, 1.1 and 2 of the *Financial Administration Act*.

Manner in which research supports government or departmental priorities: The study will assess a broad range of issues that affect quantitative public opinion research, for both telephone and online.

Manner in which research findings will benefit Canadians: The findings will help PORD to revise existing standards and guidelines and create new standards and guidelines to guide the quality of research undertaken on behalf of the Government of Canada

Method

Panel Composition

The Advisory Panel on Government of Canada Quantitative Public Opinion Research consisted of 10 members drawn from the private sector, academics experienced with market research, and Statistics Canada. Members of the Panel were recruited by PORD, with assistance from Sage Research.

Advisory Panel members gave their personal opinions and experiences on the issues discussed, and were not speaking on behalf of their organization or industry.

Private Sector	**Lorne Bozinoff** *President and CEO* **Forum Research** **Tony Coulson** *Vice-President, Corporate and Public Affairs* **Environics Research** **Raymond Cyr** *CEO* **Voxco** **Susan Galley** *Senior Vice President* **Ekos Research Associates** **Peter MacIntosh** *Executive Vice President* **Corporate Research Associates** **Rick Nadeau** *President* **Quorus Consulting Group**

	Michael Williams *President* **Advanis Inc.**
Academia	**Scott Bennett** *Associate Professor, Department of Political Science, Faculty of Public Affairs* **Carleton University** *Managing Director* **Community Analytics Institute** **Peter Loewen** *Associate Professor, Department of Political Science* *Director of the School of Public Policy and Governance* **University of Toronto**
Statistics Canada	**Pierre Caron** *Assistant Director, Methodology Branch*

Research Approach

Anita Pollak and Rick Robson of Sage Research Corporation facilitated the proceedings of the Advisory Panel. Their roles included:

1. preparation of background materials for Panel members,
2. development of discussion agendas,
3. development of proposed guidance for comment by the Advisory Panel,
4. management and facilitation of the online discussion boards and web conferences used for Advisory Panel meetings, preparation of four Working Reports,
5. and preparation of the Final Report.

The Advisory Panel was provided with copies of the current Government of Canada Standards for the conduct of telephone and online POR, and with additional background materials pertinent to the Panel's deliberations. Panelists were also provided with a literature review previously prepared for PORD (*Literature Review in Support of the 2017 Quantitative Standards* Review) which contained discussion of some of the topics the Panel was asked to address.

The Advisory Panel's work took place between April 11, 2018 and August 22, 2018.

The Advisory Panel process consisted of an initial web conference followed by three online discussion boards. Panel members then reviewed four Working Reports summarizing the results and proposing the guidance to put into the final report of the Advisory Panel.

An online discussion board method was selected to allow Panelists to log on and participate in the discussions at their convenience. For each discussion board, the moderators provided background information to the Panelists and posted questions online. Each Panelist logged on multiple times over the course of the discussion board:

(a) to post their responses to the questions from the moderators,
(b) to review and post responses or questions to the views and opinions expressed by other Panel members, and
(c) to respond to additional questions posted by the moderators throughout the discussion.

The dates of each component are shown below. Note that Sage Research informally extended the end dates for most discussion boards to allow certain Panel members needing extra time to provide their input.

Kick-off Web conference	April 11,2018
Discussion Board #1	April 16-May 4, 2018
Discussion Board #2	May 14-June 1, 2018
Discussion Board #3	June 4-June 15

The Advisory Panel's recommended guidance for quantitative research is expressed as standards and guidelines, together with supporting commentary.

Standards	Practices that are ***requirements*** for all online studies conducted by the Government of Canada; these are typically stated using the word "must."
Guidelines	Practices that are ***recommended***, but would not be requirements – that is, known good practices or criteria that serve as a checklist to ensure quality research but are not necessarily applied to every study; these are typically stated using the word "should."

While it was not a mandate of the Advisory Panel to reach consensus, it did so on quite a few aspects of standards and guidelines for quantitative research.

Panel members were provided for review and comment four Working Reports summarizing the results and proposing the guidance to put into the final report of the Advisory Panel.

Acronyms Used in the Report

AAPOR	American Association for Public Opinion Research
CRTC	Canadian Radio-television and Telecommunications Commission
ESOMAR	European Society for Opinion and Marketing Research
GC	Government of Canada
GRBN	Global Business Research Network
ISO	International Organization of Standardization
MRIA	Marketing Research and Intelligence Association[1]
MRS	Market Research Society
POR	Public opinion research
PORD	Public Opinion Research Directorate
TBS	Treasury Board Secretariat
WAI	Web Accessibility Initiative

[1] The MRIA ceased to exist on July 31, 2018. There is currently no Canadian association representing the marketing research industry.

Results

Sampling

Definitions of Types of Samples

Background and Questions

Section 4 of the Standards is *Sampling Procedures*, and it includes standards for both probability sampling and non-probability sampling. Section 4 does not give definitions of these two types of sampling procedures. This section may be revised to include definitions and some examples to help users determine whether a sample is a probability sample or a non-probability sample.

Probability sampling: Statistics Canada[2] states:

> Probability sampling is a method of sampling that allows inferences to be made about the population based on observations from a sample. In order to be able to make inferences, the sample should not be subject to selection bias. Probability sampling avoids this bias by randomly selecting units from the population (using a computer or table of random numbers). It is important to note that random does not mean arbitrary. In particular, the interviewers do not arbitrarily choose respondents since then sampling would be subject to their personal biases. Random means that selection is unbiased – it is based on chance. With probability sampling, it is never left up to the discretion of the interviewer to subjectively decide who should be sampled.
>
> There are two main criteria for probability sampling: one is that the units be randomly selected, the second is that all units in the survey population have a non-zero inclusion probability in the sample and that these probabilities can be calculated. It is not necessary for all units to have the same inclusion probability, indeed, in most complex surveys, the inclusion probability varies from unit to unit. (p.91)

Non-probability sampling: Some excerpts from the same source as above:

> Non-probability sampling is a method of selecting units from a population using a subjective (i.e., nonrandom) method. Since non-probability sampling does not require a complete survey frame, it is a fast, easy and inexpensive way of obtaining data. The problem with non-probability sampling is that it is unclear whether or not it is possible to generalize the results from the sample to the population. The reason for this is that the selection of units from the population for a non-probability sample can result in large biases.
>
> Due to selection bias and (usually) the absence of a frame, an individual's inclusion probability cannot be calculated for non-probability samples, so there is no way of producing reliable estimates or estimates of their sampling error. In order to make inferences about the population, it is necessary to assume that the sample is representative of the population. This usually requires assuming that the characteristics of the population follow some model or are evenly or randomly distributed over the population. This is often dangerous due to the difficulty of assessing whether or not these assumptions hold.

[2] Statistics Canada, 2003, *Survey Methods and Practices*, Catalogue no. 12-587-X

Questions Addressed by the Panel

The Panel was asked to comment on the following proposed expansion of Section 4.1.2, giving definitions and examples of three types of samples: probability, non-probability and census.

4. SAMPLING PROCEDURES

4.1. General [from Online Standards]

All researchers must:

1) Clearly state the target group (universe) definition for the research study; in the case of Internet surveys this includes explicit identification of whether or not non-Internet users are part of the target group definition;
2) Clearly state the method(s) used to obtain a sample of this target group, including whether the method is a probability survey, a non-probability survey, or a census. Definitions and examples of each method are as follows:
   a) Probability sample: respondents are randomly selected from the survey's target population, and each respondent's probability of inclusion can be calculated. Probability sampling is a method for obtaining a sample projectable to the target population.

      Some examples:
      - Random-digit-dial (RDD) telephone survey of Canadians.
      - Random sampling from a list of all members of the target population.
      - Random sampling from a panel that is itself a probability sample of the target population.
      - Website intercept survey in which target population is visitors to the website, and visitors are randomly sampled to take part in a survey.

   b) Non-probability sample: a sample that does not meet the requirements of a probability sample – that is, respondents are not randomly selected from the survey's target population, and/or each respondent's probability of inclusion cannot be calculated. Additional steps must be taken to try to make results from a non-probability sample representative of the target population.

      Some examples:
      - A sample drawn from a research panel consisting of people who volunteer to join the panel and do surveys. Note that a sample collected using probabilistic methods from sampling frames that were compiled using non-probability methods is considered a non-probability sample.
      - Quota sampling, in which the selection of respondents is based on judgment, convenience or some other nonrandom process.

   c) Census: An attempt is made to collect data from every member of the target population. Note that a census can be subject to other types of survey error, notably coverage error and non-response, so not every member of the target population may be in the final data set.

Advisory Panel Response

Section 4.1.2 should be expanded to include definitions and examples, and the Panel largely agreed with the proposed text.

The most contentious sentence is in 4.1.2b pertaining to non-probability surveys: *Additional steps must be taken to try to make results from a non-probability sample representative of the target population.* There were differing points of view on to what extent it might be possible to make results from a non-probability survey representative of the target population, and on whether such attempts should be required. The final decision was to drop this sentence from 4.1.2b simply because it is outside the scope of providing definitions and examples. The topic, however, continued to be considered by the Panel in the context of another topic, *Maximizing Representativeness of Non-probability Surveys***.**

Other revisions to the proposed text are:

- Replace *target group* with *target population*.
- The requirements for a probability sample have been revised to more closely match the Statistics Canada language. This is because the proposed description does not make clear that at least in principle there is a potential listing of the population from which selection can be made.
- In the reference to *research panels*, add a reference to *opt-in*.
- The description of quota sampling was unclear, and is replaced with, *Quota sampling in which respondents are selected through some nonrandom process.*
- Consider replacing *census* with *attempted census*. Note that the word *attempt* is used in the proposed definition, and the definition notes that a census is subject to certain types of survey error. The issue is whether or not to incorporate the word into the name of this sample type.

The following is the suggested revision to Section 4.1 (as applied to the Online Standards):

4. SAMPLING PROCEDURES

4.1. General [applied to Online Standards]

All researchers must:

1) Clearly state the target population (universe) definition for the research study; in the case of Internet surveys this includes explicit identification of whether or not non-Internet users are part of the target population definition;
2) Clearly state the method(s) used to obtain a sample of this target population, including whether the method is a probability survey, a non-probability survey, or a census. Definitions and examples of each method are as follows:
   a) Probability sample: The sample meets both of the following conditions: (1) respondents are randomly selected from the survey's target population, and (2) all units in the target population have a non-zero probability of being included in the sample and these probabilities can be calculated.

      Some examples:
      - Random-digit-dial (RDD) telephone survey of Canadians.
      - Random sampling from a list of all members of the target population.
      - Website intercept survey in which target population is visitors to the website, and visitors are randomly sampled to take part in a survey.

   b) Non-probability sample: a sample that does not meet the requirements of a probability sample – that is, respondents are not randomly selected from the survey's target population, and/or each respondent's probability of inclusion cannot be calculated.

Some examples:

- A sample drawn from a research panel consisting of people who volunteer to join or opt in to the panel and do surveys. Note that a sample collected using probabilistic methods from sampling frames that were compiled using non-probability methods is considered a non-probability sample.
- Quota sampling, in which respondents are selected through some nonrandom process.

c) Census: An attempt is made to collect data from every member of the target population. Note that a census can be subject to other types of survey error, notably coverage error and non-response, so not every member of the target population may be in the final data set.

Maximizing Representativeness of Non-probability Surveys

Background and Questions

Surveys based on non-probability sampling have become more common in marketing research, particularly because of the growth of online opt-in panels that provide significant cost savings over telephone probability samples. Public opinion research surveys for the Government of Canada have historically usually used probability sampling, but there may be more usage of non-probability surveys if there is confidence that these can deliver results that are representative of the target population being surveyed.

The objective is to revise the Standards to emphasize the importance of striving for representativeness in non-probability surveys, and to explain in the proposal how this will be done.

The current standards address this objective to some extent, but the intent is to make the requirement more explicit and detailed.

Current Standards

Standard 1 *Proposal Documentation*, does not contain any explicit language on the importance of maximizing representativeness. Standard 1.2.2 *Sample/Sampling Details* only says:

- State if census, probability or non-probability
- If non-probability, then *provide rationale for choosing a non-probability sample*
- 1.2.2.5 makes a reference to "quota controls" – *Define respondent eligibility/screening criteria, including any quota controls*

There is some relevant language in Standards 4.3 *Non-Probability Sampling* and 4.4 *Quota Sampling*

Standard 4.3.2 *Sampling for Non-probability Samples*

1) As for probability sampling, the list or sample source must be stated, including its limitations in covering the universe for the target sample.
2) The precise quota control targets and screening criteria must also be stated including the source of such targets (e.g., census data or other data source).

Standard 4.4 *Quota Sampling*

Quota sampling techniques are typically used for panel surveys and personal intercept studies to achieve sample representativeness. Quotas may also be used to control representativeness on other data collection methodologies.

1) A full description of the regional, demographic or other classification variable controls used for balancing the sample to achieve representativeness must be described.
2) The precise quota control targets and screening criteria must also be stated including the source of such targets (e.g., census data or other data source).

Pew Research recently published two studies[3] that examined factors affecting the extent to which results from a non-probability survey can be representative[4]. Some observations on their results:

- There are steps one can take that will improve the representativeness of non-probability survey results. It is also the case that some degree of bias will likely remain. There is no "magic method" that has been shown to eliminate all bias.
- The magnitude of bias can vary across different measures in the survey, and can vary substantially. This implies it is important to identify the key measures given the survey's purpose when making decisions about steps to take to make the survey results more representative. This is because a factor used to try to enhance representativeness will only be effective if it is correlated with the survey variables of interest.
- When selecting factors to enhance the representativeness of the sample or to make post hoc adjustments (e.g. weighting), it is better to avoid using only a small number of factors (e.g. avoid simple schemes such as gender, age and region).
- Incorporating nondemographic factors such as attitudes or behaviour correlated with the key survey measures of interest can improve representativeness providing there are high quality population statistics for such factors to use as benchmarks.
- The factors/variables selected to make adjustments tend to be more important for improving representativeness than "complex statistical methods" such as propensity weighting or sample matching. While the choice of statistical method can certainly matter, appropriate choice of adjustment variables is a critically important first step.

The following is an excerpt from the summary section of the 2018 Pew Research report, *For Weighting Online Opt-In Samples, What Matters Most?*

> ***For Weighting Online Opt-In Samples, What Matters most?*** *The right variables make a big difference for accuracy. Complex statistical methods, not so much.*
>
> A growing share of polling is conducted with online opt-in samples. This trend has raised some concern within the industry because, while low participation rates pose a challenge for all surveys, the online opt-in variety face additional hurdles. By definition they do not cover the more than 10% of Americans who don't use the internet. The fact that potential respondents are self-selected means that there is still substantial risk that these samples will not resemble the larger population. To compensate for these challenges,

[3] Pew Research Center, May 2016, "Evaluating Online Nonprobability Surveys."
Pew Research Center, January 2018, "For Weighting Online Opt-In Samples, What Matters Most?"

[4] Another interesting study that draws similar conclusions is the following, which compared non-probability surveys to "low-response rate" probability telephone surveys: Dutwin, D. & Buskirk, T., 2017, "Apples to Oranges or Gala versus Golden Delicious? Comparing data quality of nonprobability internet samples to low response rate probability samples", *Public Opinion Quarterly,* Vol. 81, Special Issues, 2017, pp. 213-249

researchers have employed a variety of statistical techniques, such as raking, propensity weighting and matching, to adjust samples so that they more closely match the population on a chosen set of dimensions. Researchers working with online opt-in samples must make a great many decisions when it comes to weighting. What factors should guide these decisions, and which ones are most consequential for data quality?

A new Pew Research Center study adds to the survey field's broader efforts to shed light on these questions. The study was based on over 30,000 online opt-in panel interviews conducted in June and July of 2016, with three vendors, and focuses on national (as opposed to state or local level) estimates. We evaluated three different weighting techniques, raking, propensity weighting and matching, both on their own and in combination. Every method was applied using two sets of adjustment variables: basic demographics (age, sex, race and ethnicity, education, and geographic region), and a more extensive set that included both demographics and a set of variables associated with political attitudes and engagement (voter registration, political party affiliation, ideology and identification as an evangelical Christian). Each procedure was performed on simulated samples ranging in size from n=2,000 to n=8,000.

The procedures were primarily appraised according to how well they reduced bias on estimates from 24 benchmark questions drawn from high-quality federal surveys. They were also compared in terms of the variability of weighted estimates, accuracy among demographic subgroups, and their effect on a number of attitudinal measures of public opinion.

Among the key findings:

- **Even the most effective adjustment procedures were unable to remove most of the bias.** The study tested a variety of elaborate weighting adjustments to online opt-in surveys with sample sizes as large as 8,000 interviews. Across all of these scenarios, none of the evaluated procedures reduced the average estimated bias across 24 benchmarks below 6 percentage points – down from 8.4 points unweighted. This means that even the most effective adjustment strategy was only able to remove about 30% of the original bias.
- **When it comes to accuracy, choosing the right variables for weighting is more important than choosing the right statistical method.** Adding a set of politically focused variables to the weighting adjustment reduced the average estimated bias by an additional 1.4 percentage points relative to adjusting only on basic demographics (e.g., age, education, race). While that might seem small, a difference of 1.4 points in the average implies that about 36 percentage points of bias were removed overall, but spread out across all 24 variables. Benchmarks most strongly associated with the political adjustment variables saw the largest improvements. In contrast, the use of more complex statistical methods never reduced the average estimated bias by more than 0.3 points beyond what was achieved with raking, the most basic statistical method evaluated.
- **The benefits of adding political variables to adjustment differ by survey topic[5].** Perhaps not surprisingly, benchmarks related to political engagement saw the largest improvement with the addition of political adjustment variables. Unweighted, these benchmarks had an average estimated bias of 22.3 percentage points, more than any other topic. While demographic weighting reduced the average bias by an average of 2.9 points, the effect of adding political adjustment variables was four times as large, reducing bias by 11.7 points and cutting the average estimated bias nearly in half (to 10.6 percentage points). Benchmarks pertaining to civic engagement and technology use also benefited disproportionately from political adjustment variables, though to a lesser degree. For benchmarks related to family composition and other personal characteristics, variable selection made little difference and proved mildly detrimental for questions of personal finance.
- **The most basic weighting method (raking) performs nearly as well as more elaborate techniques based on matching.** When weighting on both demographic and political variables, methods based on matching resulted in the lowest average bias across the full set of 24 benchmarks – either in combination with raking at smaller sample sizes (n=less than 4,000) or on its own when the sample size was larger. Even so, procedures that only used raking (the least complex method evaluated) performed nearly as well, coming in 0.1 to 0.3 points behind the most effective method,

[5] Because of the nature of many of Pew Research's interests, "political variables" are relevant to their work. The key take-away here is that appropriately chosen attitudinal or behavioural measures can help reduce bias. The nature of these variables, including whether or not they are "political", will depend on the survey topic.

depending on sample size. For benchmarks related to political engagement, the benefits from the more complex approach are somewhat larger than for other topics, doing between 0.5 and 1.2 points better than raking depending on sample size, but nowhere near the magnitude of improvement derived from weighting on political variables in addition to demographics. If the data necessary to perform matching are readily available and the process can be made routine, then a combination of matching and other methods like raking is likely worthwhile, providing incremental but real improvements. In other situations, such marginal improvements may not be worth the additional statistical labor.

- **Very large sample sizes do not fix the shortcomings of online opt-in samples.** While an online opt-in survey with 8,000 interviews may sound more impressive than one with 2,000, this study finds virtually no difference in accuracy. When adjusting on both demographic and political variables, the most effective procedure at n=8,000 was only 0.2 points better than the most effective procedure at n=2,000. While a large sample size may reduce the variability of estimates (i.e., the modeled margin of error), this is of little help from a "total survey error" perspective. For example, raking on demographic and political variables, the average modeled margin of error across all 24 benchmark variables is ±1.8 percentage points when n=2,000 and ±0.5 points when n=8,000, but the average bias holds steady at 6.3 points. As the sample size increases, estimates become less dispersed and more tightly clustered, but they are often more tightly clustered around the wrong (biased) value.

The weighting procedures tested in this report represent only a small fraction of the many possible approaches to weighting opt-in survey data. There are a host of different ways to implement matching and propensity weighting, as well as a variety of similar alternatives to raking (collectively known as calibration methods). We also did not evaluate methods such as multilevel regression and poststratification, which require a separate statistical model for every outcome variable. Add to this the innumerable combinations of variables that could be used in place of those examined here, and it is clear that there is no shortage of alternative protocols that might have produced different results.

But whatever method one might use, successfully correcting bias in opt-in samples requires having the right adjustment variables. What's more, for at least many of the topics examined here, the "right" adjustment variables include more than the standard set of core demographics. While there can be real, if incremental, benefits from using more sophisticated methods in producing survey estimates, the fact that there was virtually no differentiation between the methods when only demographics were used implies that the use of such methods should not be taken as an indicator of survey accuracy in and of itself. A careful consideration of the factors that differentiate the sample from the population and their association with the survey topic is far more important.

Questions Addressed by the Panel

The Panel was asked to comment on the following proposed revisions to Section 1, *Proposal Documentation,* and Section 4, *Sampling*.

Proposed revision to Section 1.2.2: *Proposal Documentation, Sample/Sampling Details*:

1.2.2. Sample/Sampling Details

1) Provide details related to target population:
   a) the definition of the target population in terms of its specific characteristics and geographic scope, including the assumed incidence of the population and any key subgroups and how the incidence was determined/obtained (e.g., supplied by the client);
   b) the total sample size and the sample sizes of any key subgroups.
2) Describe the sampling procedures, including:
   a) the sample source;
   b) the sample frame;
   c) whether a sample survey or a census will be conducted and, if a sample, whether probability or non-probability sampling will be applied (see section 4 for additional information to include in the proposal).

3) Explain respondent selection procedures.
4) Indicate the number of re-contact attempts and explain re-contact attempt procedures.
5) Define respondent eligibility/screening criteria, including any quota controls.
6) For non-probability samples, provide the rationale for choosing a non-probability sample. If the survey results will be used to make statements about a population, steps must be taken to maximize the representativeness of the sample with respect to the target population, and these steps must be documented in the research proposal and in the survey report (see section 4.3).

Proposed revision to Section 4.3, *Non-Probability Sampling*:

Note: This proposal includes eliminating Section 4.4, *Quota Sampling*, and moving relevant content to Section 4.3. *Non-Probability Sampling:*

For reference, Section 4.4, *Quota Sampling*, is:

Quota sampling techniques are typically used for panel surveys and personal intercept studies to achieve sample representativeness. Quotas may also be used to control representativeness on other data collection methodologies.

1) A full description of the regional, demographic or other classification variable controls used for balancing the sample to achieve representativeness must be described.
2) The precise quota control targets and screening criteria must also be stated including the source of such targets (e.g., census data or other data source).
3) Deviations from target achievement must be shown in the report (i.e., actual versus target).

4.3. Non-Probability Sampling

4.3.1. Justification of Use of Non-probability Samples

1) When a choice is made to use a non-probability sample, that choice must be justified, in both the research proposal and the research report. The justification must take into account the statistical limitations in reporting on data from a non-probability sample, and limitations in generalizing the results to the target population.

4.3.2. Sampling for Non-probability Samples

1) As for probability sampling, the list or sample source must be stated, including its limitations in covering the universe for the target sample.
2) If the survey results will be used to make statements about a population, steps must be taken to maximize the representativeness of the survey results with respect to the target population, and these steps must be documented in the research proposal and in the survey report.

   These steps include:
   - Controls on sample composition to maximize representativeness, such as quota sampling
   - Weighting
3) A full description must be provided of the regional, demographic or other classification variables used to maximize the representativeness of the sample and survey results. In selecting variables, also consider their likely correlation with key survey measures (adjustment variables that are uncorrelated with survey measures will do little to improve representativeness). Behavioural or attitudinal variables can also improve representativeness, providing relevant, high quality benchmarks exist for the target population.
4) The precise quota control targets and screening criteria must also be stated including the source of such targets (e.g., census data or other high quality data source).
5) Deviations from target achievement must be shown in the report (i.e., actual versus target).

Advisory Panel Response

The following issues were discussed with respect to the proposed text, *If the survey results will be used to make statements about a population, steps must be taken to maximize the representativeness of the survey results with respect to the target population*.

If the survey results will be used to make statements about a population: The majority of Panelists accepted the premise that survey results from a non-probability survey can possibly be used to make statements about a population, providing appropriate adjustments are made to the sample and/or data. However, a few Panelists did not agree with this premise, because they did not think that adequate adjustment methods exist to allow making statements about a population based on a non-probability sample. Under this minority view, this premise should not appear in the Standards. This would mean:

- In the text suggested above for *Proposal Documentation* 1.2.2.6, remove the premise, so that this would only say: *For non-probability samples, provide the rationale for choosing a non-probability sample.*
- In the suggested text for *4.3 Non-Probability Sampling*, remove 4.3.2.2, the section where this premise is stated.

Change "maximize the representativeness" to "improve the extent to which the survey results are representative": Among the Panelists (a majority) who accepted the premise that survey results from a non-probability survey can possibly be used to make statements about a population, this change was suggested because the steps that would be required to "maximize" representativeness may not be known with any certainty, and even if the steps are credibly known they may be of a complexity that puts the particular POR project over budget. "Improving representativeness" is more practical and realistic than "maximizing representativeness."

"Must be taken" versus "may be taken": The Panel was split on whether taking steps to improve representativeness should be worded as a requirement ("must") or a guideline ("may" or "should").

The reasons for making this a guideline rather than a requirement were (a) for some POR surveys, it may be more practical to identify the ways in which the results are not representative and to interpret the results accordingly than to try to correct for these shortcomings, and (b) there may be uncertainty at the proposal stage about what steps could be taken that would improve representativeness.

The reason for making this a requirement was that given the premise, *If the survey results will be used to make statements about a population*, it is reasonable to require that steps be taken to improve representativeness.

Option to explain not taking steps to improve representativeness: Regardless of whether taking steps to improve representativeness is worded as a requirement or a guideline, there must be an option to explain why for a particular project it may be decided not to commit to steps to improve representativeness. This decision might be made because at the proposal stage there is uncertainty about what, if any, steps can realistically be taken, or there may be other reasons.

Among the Panelists who accepted the premise that survey results from a non-probability survey can possibly be used to make statements about a population, the following are two alternative wordings that reflect the split on whether taking steps to improve representativeness should be a requirement or a guideline:

Option 1: Taking steps is a requirement

If the survey results will be used to make statements about a population, steps must be taken to improve the extent to which the survey results are representative of the target population, subject to special justifications to the contrary. These steps must be documented in the research proposal and in the survey report.

These steps could include:

- Controls on sample composition to improve representativeness, such as quota sampling
- Weighting

If at the proposal stage no specific measures to improve representativeness are planned to be taken, the reasons for this must be documented in the research proposal. If subsequently no measures are taken to improve representativeness, the reasons for this must be documented in the survey report.

Option 2: Taking steps is a guideline

If the survey results will be used to make statements about a population, steps should be taken to improve the extent to which the survey results are representative of the target population, and any steps taken must be documented in the research proposal and in the survey report.

These steps could include:

- Controls on sample composition to improve representativeness, such as quota sampling
- Weighting

If at the proposal stage no specific measures to improve representativeness are planned to be taken, the reasons for this must be documented in the research proposal. If subsequently no measures are taken to improve representativeness, the reasons for this must be documented in the survey report.

There were some comments on the proposed text in 4.3.2, *In selecting variables, also consider their likely correlation with key survey measures (adjustment variables that are uncorrelated with survey measures will do little to improve representativeness). Behavioural or attitudinal variables can also improve representativeness, providing relevant, high quality benchmarks exist for the target population."* One suggestion was to change *"high quality benchmarks"* to *"population statistics"* on the grounds that the latter is simpler and more direct about what is required. Some other comments did not involve changes to the proposed text, but rather are observations on difficulties that can arise:

- It may not be clear what would be good adjustment variables, particularly if there is no past research on the topic.
- When "high quality benchmarks" (or "population statistics") exist, they may not be recent, raising a question of how old the data can be and still be valid for adjustment purposes.
- In practice, it is often a struggle to limit the number of questions in a survey to a reasonable duration. In this context, adding behavioural or attitudinal variables with an uncertain impact on improving representativeness may be hard to justify.

One Panelist suggested adding a requirement or guideline when a non-probability sample is being used to show that the proposed approach is based on a methodology that has produced externally validated estimates in the past. The Panelist commented that there has been some progress on developing approaches to produce validated results in specific areas such as voting intention. Another Panelist, however, argued that this would not be feasible as a requirement, because:

- Often the time available to prepare a proposal is quite short, and there is not sufficient time to research the performance of the proposed methodology.
- Other than voting intention surveys, there may be little or no previous research with the same proposed methodology that can be used to assess performance in producing validated estimates.

Online Sample Information to Include in *Proposal Documentation*

The current standards contain the following requirements for proposal documentation when an online sample provider is used:

1.2.4. Description of Data Collection

3) For access panels, a description of the following must be provided, at minimum (when multiple panels are used, the information must be provided for each):
    a) active panel size (provide the definition of "active");
    b) panel recruitment;
    c) panel monitoring;
    d) panel maintenance;
    e) panel refreshment.

The objective is to update and revise this standard, including:

- Distinguish between the information requirements for probability-based panels versus non-probability panels
- Recognize that river sampling may be used as a sample source
- Provide clarity as to what specific information is to be included in the proposal
- Provide guidance, if possible, on good practices the client should look for

Note that there are other relevant sections in the Standards, notably:

- Section 4 *Sampling Procedures*
- Section 5.1.5 *Use of respondent lists*
- Section 7.5 (Online Standards) *Access Panel Participation Frequency*
- Section 1.2.4. #5: *Describe the planned fieldwork validation methods and procedures*

The intent is to not duplicate standards in these areas. Therefore, the focus here was specifically on additional proposal documentation requirements when an online sample provider is used who will be providing sample using an access panel or river sampling.

Questions Addressed by the Panel

- What should be the proposal information requirements for a probability online panel sample?
- What does it mean for a panel to be "large enough" for purposes of fielding a particular survey?
- What should be the proposal information requirements for a non-probability online sample?

Background and Questions: What Does it Mean for an Online Panel to be "Large Enough"?

For both probability and non-probability surveys in the revised Section 1.2.4 #3/#4, it says the following with respect to providing information in the proposal about panel size:

> - Active panel size (provide the definition of "active"), and if possible the active panel size corresponding to the survey's target population that is available to be surveyed.
>
> The panel size should be large enough to ensure a high likelihood the target sample size of usable completed questionnaires can be obtained, taking into account such factors as likely response rate, exclusion rules involving past survey participation, and sample reduction due to data cleaning.

PORD asked that the Panel comment on how to assess *"large enough."*

Advisory Panel Response

There is no numeric rule of thumb that can be given in the Standards to assess what constitutes "large enough." There are a variety of project-specific factors that can affect how large the sample of survey invitations needs to be in order to deliver the desired number of completed questionnaires. Factors include incidence rate, response rate, fieldwork duration, exclusion rules involving past survey participation, and data loss resulting from data cleaning. Response rate is itself influenced by a variety of factors such as survey topic, sponsor, questionnaire length, and incentive. Ultimately, there is quite a bit of judgment on the part of the research supplier to estimate the required number of survey invitations for any particular project.

The Panel was split on whether or not there should be a requirement to report the active panel size:

- Some Panelists supported the suggested text above for Section 1.2.4 #3/#4. The role of the paragraph referring to *"large enough"* is simply to flag that thought needs to be given to various factors that will affect the required number of survey invitations, and that this needs to be checked against the size of the panel.
- Some other Panelists suggested dropping the requirement to state active panel size in the research proposal. The rationale is (a) this is not actionable information for the research client because ultimately they are reliant on the research supplier's judgement about whether the panel is large enough for the project, and (b) a core job of the research supplier is to estimate the required number of survey invitations, and they do not need to be reminded about this. From this perspective, informing the client about the panel size only becomes relevant if the panel is not large enough and as a result either the survey sample size has to be revised or more

than one panel needs to be used. These circumstances will trigger informing the client of panel size based on other existing and proposed standards.

Background and Questions: Proposal Documentation Requirements for a Probability Online Panel Sample

For purposes of discussion by the Panel, the current Section 1.2.4 #3 was substantially revised, including suggesting separate proposal information requirements for probability versus non-probability samples.

The proposed requirements for probability online panel samples were:

1.2.4. Description of Data Collection

3) When the sample is a **probability sample** of the survey's target population drawn from a probability-based panel, a description of the following must be provided, at minimum (when multiple panels are used, information must be provided for each):

- Active panel size (provide the definition of "active"), and if possible the active panel size corresponding to the survey's target population that is available to be surveyed

  The panel size should be large enough to ensure a high likelihood the target sample size of usable completed questionnaires can be obtained, taking into account such factors as likely response rate, exclusion rules involving past survey participation, and sample reduction due to data cleaning.

- How the panel is constructed and maintained such that it is a probability sample
- How a probability sample of the survey's target population will be drawn from the panel
- Whether or not respondents will be sourced from more than one panel. If more than one panel is being used, provide the following:
  - -- The identity of the other panels
  - -- Whether or not each of the other panels is a probability-based panel

    If the sample includes respondents from non-probability samples, then the sample as a whole is a non-probability sample.
  - -- The reasons for sourcing sample from more than one panel (e.g. to achieve the desired survey sample size; to get better coverage of the survey population)
  - -- Steps to be taken to ensure that people enrolled in more than one of the panels are not sent the survey more than once
- The incentive or reward respondents are offered for taking part in the survey

  This should be reviewed to ensure it does not risk causing certain types of people to either take or avoid the survey, and thereby potentially bias the sample.

Advisory Panel Response

Given that there are now online panels that purport to be probability samples, the Panel agreed that there needs to be proposal documentation requirements specific to such samples.

The Panel agreed with the proposed text for the information that needs to be in a POR research proposal for a survey using a probability online panel sample.

In this regard, the single most important required item is *How the panel is constructed and maintained such that it is a probability sample*. The majority of Panelists were either skeptical that it is actually possible to have probability online panel, or skeptical that it can be as good as a RDD probability sample. The skepticism is based on the characteristics of an opt-in research panel. The people who are willing to join and actively participate in an online opt-in panel over time may be different in some ways from those who are not willing to participate, even if panelists are initially recruited to the panel using probability sampling. Actively participating in a research panel typically means responding to a large number of surveys over time. This high degree of involvement in and exposure to surveys can potentially affect in some ways how panelists respond to survey questionnaires. The requirement to document *How the panel is constructed and maintained such that it is a probability sample* means any concerns are addressed in the POR proposal.

Background and Questions: Proposal Documentation Requirements for Non-probability Samples from an Online Sample Provider

The objective is to update and revise *Proposal Documentation* Standard 1.2.4 #3 for non-probability samples obtained from an online sample provider. Considerations include:

- Recognize that river sampling may be used as a sample source
- Provide clarity as to what specific information is to be included
- Provide guidance, if possible, on good practices the client should look for

For discussion purposes, the Panel was provided with three different listings of what the required proposal documentation might be:

- The current *Proposal Documentation* 1.2.4 #3
- ESOMAR *Online Research*, Section 6.1, *Sample Source and Management*
- Proposed revision above in the Background

The following pages show the guidance from the three sources.

1.2.4. Description of Data Collection – Alternative approaches

3) For access panels, a description of the following must be provided, at minimum (when multiple panels are used, the information must be provided for each):

Current 1.2.4	**ESOMAR *Online Research* (pp. 18-19)**	**A proposed revision to 1.2.4 #4 (the non-probability section)**
• active panel size (provide the definition of "active"); • panel recruitment; • panel monitoring; • panel maintenance; • panel refreshment.	• details of how the sample was recruited and a description of the sampling frame and how well the sample covers the target population it is meant to represent • steps taken to validate sample sources • the procedures used to "on-board" prospective participants to panels • cleaning and updating procedures • any monitoring of survey-taking performance or quality controls to minimize satisficing or fraud and the steps taken if such behaviour is identified** • whether and how new sources are integrated • any procedures in place to maximize sample consistency for tracking projects	When the sample is a non-probability sample drawn from a panel or other online sources and obtained from an online sample provider, a description of the following must be provided, at minimum (when multiple panels are used, information must be provided for each). Note that "panel" refers both to panels operated by an online sample provider and to lists available from online sample providers. • Whether all or part of the sample will come from web-intercept methods ("river sampling") or other non-panel sources, and if so what these sources are and how respondents will be selected for your survey These methods may involve less quality control in terms of validation of respondent identity. The nature of the websites from which respondents are sourced can potentially introduce bias: a large number of diverse websites is better than a small number of websites. • If a panel is being used, the active panel size (provide the definition of "active"), and if possible the active panel size corresponding to the survey's target population The panel size should be large enough to ensure a high likelihood the target sample size of usable completed questionnaires can be obtained, taking into account such factors as likely response rate, exclusion rules involving past survey participation, and sample reduction due to data cleaning. • If a panel is being used, whether or not respondents will be sourced from more than one panel. If more than one panel is being used, provide the following: -- The identity of the other panels. -- The reasons for sourcing sample from more than one panel (e.g. to achieve the desired survey sample size; to get better coverage of the survey population; to reduce the risk of bias) -- Steps to be taken to ensure that people enrolled in more than one of the panels are not sent the survey more than once.

		• If a panel is being used, describe the recruitment sources used to populate the panel on an ongoing basis, and if possible information on the proportion of panelists accounted for by each source. The goal is to ensure the diversity of the survey population is represented, not only in terms of demographics but also on relevant behaviours and attitudes. In this regard, it can be better to have a large number of diverse sources, to reduce the risk of bias. • If a router is used to select survey respondents, what measures are taken to guard against, or mitigate, any bias that might arise from using a router. A router is an online system that screens potential respondents and assigns each person to one of multiple surveys being fielded at the same time. Routers are important for efficiently fielding multiple surveys. However, there can be a risk that systematically directing people with certain characteristics to other surveys may bias the profile of people directed to your survey. • If profile data from a panel is used to structure the sample or for key analyses, and these measures are not validated by remeasuring the characteristics in the questionnaire, provide information on the panel's validation processes to ensure the information is accurate and up to date. Panel profile data on panelists can be very helpful for structuring a sample. However, it is a good practice to verify this data by remeasuring the characteristics in the survey questionnaire. • If the survey is a tracking study, the steps that will be taken to ensure comparability of the sample with previous waves of the tracking study • The incentive or reward respondents are offered for taking part in the survey This should be reviewed to ensure it does not risk causing certain types of people to either take or avoid the survey, and thereby potentially bias the sample.

** Section 7.8 of the Standards addresses satisficing. The question is whether this should also be referenced in Section 1, *Proposal Documentation*

7.8. Detecting and Dealing with Satisficing

1) The access panel provider, working with the client, shall implement procedures to identify and remove fraudulent and inattentive panel members or respondents, documenting these procedures and actions taken.

Advisory Panel Response

The current proposal documentation requirements for access panels in Section 1.2.4 #3 of the Online Standards are not sufficient. The current standard is:

1.2.4. Description of Data

3) For access panels, a description of the following must be provided, at minimum (when multiple panels are used, the information must be provided for each):
   a) active panel size (provide the definition of "active");
   b) panel recruitment;
   c) panel monitoring;
   d) panel maintenance;
   e) panel refreshment.

The basic problem with this standard is that it is too vague, and it is not clear what specific information is supposed to be provided. It also does not given any indication of why the information might matter or how to use it. Another issue is that not all of the information may be available, either because the panel provider considers it to be proprietary, or because of limited time for proposal preparation.

There were differing views within the Panel on what information should be required about a non-probability online sample obtained from an online sample provider.

One point of view was that only very basic information should be required in the proposal, such as whether the sample will be drawn from a panel or panels, and the identity of the panel(s). The rationale is that details about a panel are not actionable information for the client, and may not have any clear relationship to the quality of the survey sample. It is also time consuming to compile detailed information, and some of it may not be available. What is arguably more important for the client and for survey quality is what steps will be taken to improve the representativeness of the sample – and this is addressed elsewhere in the Standards.

The majority of Panelists supported having some information disclosure requirements for a non-probability online sample obtained from an online sample provider, although there were some different points of view on the specifics.

The following are possible elements for a revised Section 1.2.4 #3 of the Online Standards, with commentary.

Introductory Paragraph

1.2.4. Description of Data Collection

When the sample is a non-probability sample drawn from a panel or other online sources and obtained from an online sample provider, a description of the following must be provided, if available. Note that "panel" refers both to panels operated by an online sample provider and to lists available from online sample providers.

The phrase *"if available"* is used because some information may not be available, either because the sample provider considers it to be proprietary, or it is not available at the time the proposal is prepared.

First Bullet Point

- Whether all or part of the sample will come from web-intercept methods ("river sampling") or other non-panel sources, and if so what these sources are and how respondents will be selected for your survey.

  These methods may involve less quality control in terms of validation of respondent identity. The nature of the websites from which respondents are sourced can potentially introduce bias. If these sources are likely to lead to potential bias in the sample composition, this should be noted and any steps taken to mitigate bias should be described.

Second Bullet Point

- Active panel size (provide the definition of "active"), and if possible the active panel size corresponding to the survey's target population that is available to be surveyed.

  The panel size should be large enough to ensure a high likelihood the target sample size of usable completed questionnaires can be obtained, taking into account such factors as likely response rate, exclusion rules involving past survey participation, and sample reduction due to data cleaning.

The Panel was split on whether or not this should be a requirement. The rationale for including versus not including it is the same as was discussed earlier in the section on information disclosure requirements for probability panels.

Third Bullet Point

- If a panel is being used, the identity of the panel and whether or not respondents will be sourced from more than one panel. If it is not known at the time of proposal preparation if more than one panel will be used but the supplier deems this may be needed, this possibility should be noted in the proposal.

  If at the time of proposal preparation it has already been decided to use more than one panel, provide the following:

  -- The identity and sizes of the other panels.
  -- The reasons for sourcing sample from more than one panel (e.g. to achieve the desired survey sample size; to get better coverage of the survey population; to reduce the risk of bias).
  -- Steps to be taken to ensure that people enrolled in more than one of the panels are not sent the survey more than once.

If it is not known at the time of proposal preparation…: This qualifier is needed because it may not be known with certainty at the time of proposal preparation whether or not multiple panels will be needed. This can depend on discussions with the client after a contract has been awarded on such matters as qualified respondents, incidence, and sample sizes – both in total and for subgroups.

The identity and sizes of the other panels: As noted previously, the Panel was split on whether information on panel size should be required.

Steps to be taken to ensure that people enrolled in more than one of the panels are not sent the survey more than once: One Panelist suggested this requirement be removed because they did not think it is feasible: panels may not want to share information on their members with other panels.

Fourth Bullet Point

- If a panel is being used, describe the recruitment sources used to populate the panel on an ongoing basis, and if possible information on the proportion of panelists accounted for by each source.

  The goal is to ensure the diversity of the survey population is represented.

The Panel was split on whether or not information on recruitment sources should be required. Some thought this is useful information, while others did not. The rationale for not having this requirement is that there is already guidance in the Standards to outline any steps that will be taken to improve the representativeness of a non-probability sample. The latter is the key information for the research client, and given this, information on recruitment sources is arguably not actionable for the research client. Also, the panel provider may consider recruitment sources to be proprietary information.

Fifth Bullet Point

- If profile data from a panel is used to structure the sample or for key analyses, and these measures are not validated by remeasuring the characteristics in the questionnaire, provide information on the panel's validation processes to ensure the information is accurate and up to date.

  Panel profile data on panelists can be very helpful for structuring a sample. However, it is a good practice to verify this data by remeasuring the characteristics in the survey questionnaire.

The Panel was split on this requirement. The issue was whether or not usage of panel profile data should be allowed without remeasuring the characteristics in the survey questionnaire.

Some Panelists said it should be a requirement that the survey questionnaire include questions to verify any panel profile data being used. This would essentially result in eliminating this bullet point, and adding a requirement to this effect to Section 2, *Questionnaire Design*. This ensures that the data for these characteristics is accurate and up to date.

Some other Panelists said that using good quality panel profile data can reduce respondent burden by reducing questionnaire length. The proposed text ensures confidence in the quality of the panel profile data being used in the survey.

Sixth Bullet Point

- The incentive or reward respondents are offered for taking part in the survey

  This should be reviewed to ensure it does not risk causing certain types of people to either take or avoid the survey, and thereby potentially bias the sample.

Statistical Treatment of Survey Results

Statistical Treatment of Non-Probability Survey Results

Background and Questions

The issue here is what can and should be done with respect to the statistical treatment of non-probability surveys.

The current standards are in Section 4.3.3, *Statistical Treatment of Non-probability Samples*. Note that this section comes after standards pertaining to justification of the use of a non-probability sample (Section 4.3.1), and requirements for information disclosure and improving representativeness (Section 4.3.2).

Current Standards

4.3.3. Statistical Treatment of Non-probability Samples

1) There can be no statements made about margins of sampling error on population estimates when non-probability samples are used.
2) The survey report must contain a statement on why no margin of sampling error is reported, based on the following template: "Respondents for this survey were selected from among those who have [volunteered to participate/registered to participate] in online surveys. The results of such surveys cannot be described as statistically projectable to the target population. [If weighting was done, state the following sentence on weighting:] The data have been weighted to reflect the demographic composition of (target population). Because the sample is based on those who initially self-selected for participation [in the panel], no estimates of sampling error can be calculated."

   This statement must be prominently placed in descriptions of the methodology in the survey report, including the executive summary.
3) In general, for non-probability surveys it is not appropriate to use statistical significance tests. However, tests of significance with non-probability samples are appropriate when the objective is to establish the extent of the relationship among variables. If tests of significance are used with non-probability samples, it must be clearly noted that conclusions from these tests cannot be generalized to any population.

   Any use of descriptive statistics must clearly indicate that they are not formally generalizable to any group other than the sample studied, and there cannot be any formal statistical inferences about how the descriptive statistics for the sample represent any larger population.

   In the case of non-probability surveys that employ an experimental design in which respondents are randomly assigned to different cells, it is appropriate to use statistical significance tests to compare results from different cells.

The 4.3.3 standards are consistent with, albeit more detailed than, the Marketing Research and Intelligence Association (MRIA) standards (note that these apply to research results generally, not just non-probability samples)[6]:

> **MRIA Code of Conduct**
>
> Researchers must not present research results with greater confidence than the data warrants. Instead, as responsible professionals, members must point out the relevant limitations of the research. This includes but is not limited to the following guidelines:
>
> i. Disclosing relevant potential sources of error, both sampling and non-sampling (e.g. response, non-response, measurement, coverage, etc.).
>
> ii. Being explicit about the assumptions made about data accuracy when employing quota or stratification methods with probability samples.
>
> iii. Refraining from making unqualified statements about confidence intervals or margins of sampling error on population estimates when probability samples are not used. For example, panels of repeat volunteers will not ordinarily qualify as sources of probability samples of the general population."

Until 2015, the American Association for Public Opinion Research (AAPOR) took a similar stance. However, in 2015 AAPOR revised its *Code of Professional Ethics and Practices* to allow for reporting of measures of precision from non-probability samples.

The change was apparently motivated in part by the 2013 *Report of the AAPOR Task Force on Non-probability Sampling*, which stated:

> We believe that users of non-probability samples should be encouraged to report measures of the precision of their estimates, but suggest that, to avoid confusion, the set of terms be distinct from those currently used in probability sample surveys. The precision of estimates from non-probability samples is not the average deviation over all possible samples, but rather is a model-based measure of deviation from the population value. Ipsos, for example has proposed the credibility interval … for their estimates from an opt-in panel survey. As noted in Section 6, the credibility interval is measure of uncertainty that is used with Bayesian methods, and Ipsos described their procedure as Bayesian. Other model-based approaches also produce estimates of precision such as standard errors that could be used and do not refer to the average over all possible samples (the accepted terminology for design-based inferences used in probability samples).
>
> Although the research base does not exist to endorse this particular measure or to urge its adoption across the industry, we believe the industry needs constructive attempts to develop measures that fill the gap created by the unsuitability of the standard margin of error calculation with non-probability samples. Treating estimates as though they had no error at all is not a reasonable option. At this point, it falls to individual researchers to judge the usefulness of this particular measure. Such judgments are only possible when organizations using them fully disclose the full range of information specified in the AAPOR Code of Professional Ethics and Practice along with a detailed description of how the underlying model was specified, its assumptions validated, and the measure calculated.

The relevant section of AAPOR's Code of Professional Ethics and Practice now reads as follows:

> III. Standards for Disclosure
>
> A. Disclosure Items for Surveys

[6] MRIA Code of Conduct for Market and Social Research, Appendix L – Polling Standards for the Canadian Marketplace, June 2017

10. Sample sizes (by sampling frame if more than one was used) and a discussion of the precision of the findings.

    For probability samples, the estimates of sampling error will be reported, and the discussion will state whether or not the reported margins of sampling error or statistical analyses have been adjusted for the design effect due to weighting, clustering, or other factors.

    Disclosure requirements for non-probability samples are different because the precision of estimates from such samples is a model-based measure (rather than the average deviation from the population value over all possible samples). Reports of non-probability samples will only provide measures of precision if they are accompanied by a detailed description of how the underlying model was specified, its assumptions validated and the measure(s) calculated. To avoid confusion, it is best to avoid using the term "margin of error" or "margin of sampling error" in conjunction with non-probability samples.

AAPOR has issued detailed guidance in *AAPOR Guidance on Reporting Precision for Nonprobability Samples*:

- Guidance is reported for Bayesian credible intervals as well as some other approaches.
- The guidance also includes an example of a statement that could be included in a survey report.
- The document notes that use of one these estimates for precision for a non-probability sample is <u>not</u> required:

  > For some surveys (e.g., exploratory, internal research) estimating precision may not be important to the research goals. For other surveys precision measures may be relevant, but the researcher may not have the statistical resources to compute them. Under the AAPOR Code, it is acceptable for researchers working with nonprobability samples to decline to report an estimate of variance. In such cases, it may be useful to note that the survey estimators have variance, but there has been no attempt to quantify the size.

Note that the current Section 4.3.3 can be interpreted as not permitting alternative methods of statistical inference such as Bayesian credible intervals. This is based on the sentence in 4.3.3 #3, *Any use of descriptive statistics must clearly indicate that they are not <u>formally</u> generalizable to any group other than the sample studies, <u>and there cannot be any formal statistical inferences about how the descriptive statistics for the sample represent any larger population</u>.*

Note that the MRIA guideline does not forbid using Bayesian credible intervals, as the language used refers only to margin of sampling error. Bayesian credible intervals are not a margin of sampling error. The MRIA guidelines are silent on the use of Bayesian credible intervals.

The objective is to update and clarify Section 4.3.3 on use of statistical measures with non-probability samples.

Questions Addressed by the Panel

The Panel was asked to consider the following two alternative options with respect to possible revisions of Section 4.3.3, *Statistical Treatment of Non-probability Samples*:

- *Option 1:* Expressly forbid the use of alternative measures of precision such as Bayesian credible intervals in GC POR surveys.

- *Option 2:* Expressly permit the use of alternative measures of precision such as Bayesian credible intervals in GC POR surveys.

For purposes of discussion, the Panel considered the following revision to Section 4.3.3: *Statistical Treatment of Non-probability Samples*. The proposed text was written for *Option 2*. Note, though, that some of the revisions would apply equally to *Option 1*. The main point of difference is in 4.3.3 #3.

4.3.3. Statistical Treatment of Non-probability Samples

1) There can be no statements made about margins of sampling error on population estimates when non-probability samples are used.

   Also, there can be no statement that the sample has a level of error equivalent to that of a probability sample of similar size.

   There can be no tests of statistical significance that are based on estimates of sampling error. An exception to this is a non-probability survey that employs an experimental design in which respondents are randomly assigned to different cells. In this case it is appropriate to use statistical significance tests based on estimates of sampling error to compare results from different cells.

2) The survey report must contain a statement on why no margin of sampling error is reported, based on the following template: "Respondents for this survey were selected using non-probability sampling methods. Because of this, margins of sampling error and tests of statistical significance based on sampling error cannot be reported."

   This statement must be prominently placed in descriptions of the methodology in the survey report, including the executive summary.

3) There are alternatives for margin of sampling error for estimating precision that can be used with non-probability samples, such as Bayesian credible intervals. Researchers have a choice of whether or not to use these alternatives.
   - For some surveys (e.g., exploratory, internal research) estimating precision may not be important to the research goals. For other surveys precision measures may be relevant, but the researcher may not have the statistical resources to compute them. It is acceptable for researchers working with non-probability samples to decline to report an estimate of precision. In such cases, the report must note that survey estimators have variance, but there has been no attempt to quantify the size.
   - If an alternative statistical measure of precision such as Bayesian credible intervals is used:
     - OPTIONAL: The following alternatives are accepted:
       - Bayesian credible intervals
       - Resampling approaches
       - Taylor series linearization
     - The statistical measure of precision that will be used must be stated in the research proposal, together with a rationale and brief description.
     - The survey report must provide:
       - A detailed description of how the underlying model was specified, its assumptions validated and the measure(s) calculated. Refer to the AAPOR document *AAPOR Guidance on Reporting Precision for Nonprobability Samples* for the information to provide, as well as an example of the type of statement to make in the report.

- One key assumption is that the survey results are unbiased. This assumption must be prominently noted, together with any limitations on this assumption (see 4.3.1).
- An explanation of how to understand the measure of precision.
- If applicable, how tests of statistical significance of differences or relationships based on the alternative method are to be understood.

Advisory Panel Response

Most Panelists supported Option 1 – i.e. expressly forbidding the use of alternative measures of precision such as Bayesian credible intervals. There were two basic concerns expressed with respect to allowing use of alternative measures:

- There is not yet a consensus on standard, proven practices in terms of what constitutes appropriate statistical models and model assumptions for POR surveys. The result can be that measures of precision will vary across suppliers, and across different surveys. For a given data set, different suppliers might produce different estimates of the precision of the survey measures.
- Alternative measures of precision for non-probability surveys require special statistical expertise, and these measures are not familiar to many research users in the same way as margin of sampling error in the case of probability surveys. In this context, there was concern that it would be difficult for many POR clients and users to evaluate a supplier's use of these measures and to understand how to use them correctly when interpreting survey results.

One Panelist supported Option 2 – i.e. allowing optional use of alternative measures of precision, but with a significant condition – namely, that the non-probability survey methodology used for a study has been previously validated against observable results. This would mean there is confidence that the non-probability survey is generating reasonable population estimates, and in this context use of an alternative measure of precision is a reasonable option.

One Panelist was strongly in favour of Option 2, allowing optional use of alternative measures of precision:

- The requirements to document and explain the alternative measure will ensure that suppliers make reasonable and appropriate choices, and that research users will understand how to interpret the measure of precision.
- Not stating any measure of precision for a survey estimate will probably result in many research users ascribing more confidence to the single point estimate reported than is justified, and they may fail to understand the degree of uncertainty around that point estimate.
- Alternatives such as Bayesian credible intervals are well developed statistical methods, even though application to POR surveys has been limited and is relatively recent. Given the increasing challenges of conducting probability POR surveys, and the cost advantages that non-probability surveys can have, it is important for the industry to encourage development and understanding of these alternatives.

The majority of Panelists supported a compromise wording for a standard of the following sort: *The use of alternative measures of precision such as Bayesian credible intervals are not allowed for GC*

POR surveys unless approved by PORD and agreed to by the GC client. This opens the door to optional use of alternative measures of precision, but explicitly emphasizes the need to ensure buy-in by PORD and the GC client on a case by case basis. It also allows a GC client to include a requirement in a survey's statement of work for an alternative measure of precision such as Bayesian credible intervals if that is something they want to have. One Panelist suggested adding a further condition that any usage of alternative measures of precision must be based on a request in the client's Statement of Work. One Panelist did not support the compromise wording because they did not support any usage of alternative measures of precision for GC POR surveys.

This compromise wording would be accompanied by requirements similar to those in the proposed text given to the Panel for Section 4.3.3 #3. In this regard, the Panel suggested the following modifications to the proposed text:

- Do not list what alternative measures of precision are acceptable, because methods are evolving.
- Do not refer specifically to the AAPOR guidance in this area. If any of this guidance is judged useful for the Standards, it should be stated directly in the Standards.

With these modifications, the majority of Panelists supported the following revised Section 4.3.3 #3:

4.3.3. Statistical Treatment of Non-probability Samples

3) The use of alternative measures of precision such as Bayesian credible intervals are not allowed for GC POR surveys unless approved by PORD and agreed to by the GC client.
   - If a measure of statistical precision is not used, the report must note that survey estimators have variance, but there has been no attempt to quantify the size.
   - If an alternative statistical measure of precision such as Bayesian credible intervals is used:
     - The statistical measure of precision that will be used must be stated in the research proposal, together with a rationale and brief description.
     - The survey report must include the following:
       - A detailed description of how the underlying model was specified, its assumptions validated and the measure(s) calculated. This must include any model assumption about the survey results being unbiased and any limitations on this assumption.
       - An explanation of how to understand the measure of precision, and any limitations or cautions in interpreting the measure.
       - If applicable, how tests of statistical significance of differences or relationships based on the alternative method are to be understood.

The text proposed to the Panel for Section 4.3.3 #1 and #2 was accepted as follows:

4.3.3. Statistical Treatment of Non-probability Samples

1) There can be no statements made about margins of sampling error on population estimates when non-probability samples are used.

   Also, there can be no statement that the sample has a level of error equivalent to that of a probability sample of similar size.

   There can be no tests of statistical significance that are based on estimates of sampling error. An exception to this is a non-probability survey that employs an experimental design

in which respondents are randomly assigned to different cells. In this case it is appropriate to use statistical significance tests based on estimates of sampling error to compare results from different cells.

2) The survey report must contain a statement on why no margin of sampling error is reported, based on the following template: "Respondents for this survey were selected using non-probability sampling methods. Because of this, margins of sampling error and tests of statistical significance based on sampling error cannot be reported."

   This statement must be prominently placed in descriptions of the methodology in the survey report, including the executive summary.

Statistical Treatment of Probability Survey Results

Background and Questions

PORD asked the Panel to consider the points made in a short article by Robert Peterson with the provocative title, *It's time for pollsters to report margins of error more honestly*[7]. The article lists the following problems with current practices:

> First, the reported margin of sampling error may imply that it pertains to an entire poll. In reality, it is only for a single question, even though research shows that many of us think a reported margin pertains to the entire poll.
>
> Second, the single question used to calculate the margin is a fictitious one. It does not reflect the actual response to any question posed in the poll.
>
> And finally, the formula for calculating the margin arbitrarily assumes that the answer to the fictitious question is 50 percent "Yes" and 50 percent "No." Only if, by chance, were the polled individuals to respond exactly that way would the reported margin be correct for that question.
>
> The rationale offered for reporting margins of error this way is that a 50/50 split has the largest possible margin. If a poll reports the maximum sampling error possible, it is covering the worst-case scenario — for that (fictitious) question.
>
> Except that the poll hasn't necessarily done so. If a question has more than two possible answers — like, "Don't Know" in addition to "Yes" and "No" — it requires a different formula. For subgroups such as "males" or "65 and over," the margins will be larger than that for the overall poll.
>
> Most important, margins of error can be cumulative. Many answers to questions in a poll are related to one another. Collectively, their margins add up to more than that for a single question. In a simple poll with only four questions, and a 95 percent confidence rate for each one, the total margin of sampling error of the poll might be as large as 19 percent.

The article lists the following types of solutions to these problems:

> There are more accurate ways for polls to represent their margins of error. One option is to present none at all. A recent Harvard CAPS-Harris Poll did just that.
>
> Another option is to compute the individual sampling error margins for every question in a poll. Pollsters could then work out an average margin across all poll questions. Such

[7] Peterson, R. (2018). *It's time for pollsters to report margins of error more honestly*. University of Texas at Austin, *UT News*, March 1, 2018.

estimates would still possess weaknesses, but they would still be superior to the single fictitious margin currently being reported.

Ultimately, pollsters should devise a multiquestion approach. Such an approach would simultaneously take into account the number, types, weights and relationships of questions, along with individual sampling error margins for actual answers and one for the overall poll.

Section 14.7 (online)/15.7 (telephone) addresses one of the points in the article, namely that margin of sampling error varies by sample size – i.e. the margin of sampling error will be different for subgroups versus the total sample:

14.7/15.7. Mandatory Survey Report Requirements; Results

2) For probability samples, state the level of precision, including the margin of error and confidence interval for the total sample and any key subgroups.

Questions Addressed by the Panel

For discussion purposes, the Panel was provided with a proposed revision to 14.7.2/15.7.2 that adds a requirement to indicate how the margin of sampling error varies across different observed percentages.

Revised 14.7/15.7. Mandatory Survey Report Requirements; Results

2) For probability samples, state the level of precision, including the margin of sampling error and confidence interval for the total sample and any key subgroups, and for a selection of different percentage values spanning the range of percentages that appear in the report.

The Panel was asked to comment on this proposed revision, and on any other aspects of the Peterson article that might impact the Standards.

Advisory Panel Response

The Panel was split on what the requirements should be for reporting level of precision for probability surveys.

Each of the following two options for 14.7/15.7 was supported by several Panelists:

- *Option 1:*

  For probability samples, state the level of precision, including the margin of sampling error and confidence interval for an observed percentage of 50% for the total sample and any key subgroups.

  Results for subgroups will have a larger margin of sampling error because of their smaller sample sizes. Also, the margin of sampling error is highest for questions where 50% of the respondents gave one answer and the other 50% gave another answer. The margin of sampling error decreases as the observed percentage for a particular response approaches 0% or 100%.

  Applying this standard means that numeric margins of sampling error would be reported for an observed percentage of 50% on the base of the total sample, and on the bases of key subgroups.

One Panelist suggested the survey report should also provide a link to a website with a calculator for readers who want to calculate margins of sampling error for other observed percentages or sample sizes.

- *Option 2:*

  For probability samples, state the level of precision, including the margin of sampling error and confidence interval for the total sample and any key subgroups, and for a selection of different percentage values representative of the range of percentages that appear in the report.

  Results for subgroups will have a larger margin of sampling error because of their smaller sample sizes. Also, the margin of sampling error is highest for questions where 50% of the respondents gave one answer and the other 50% gave another answer. The margin of sampling error decreases as the observed percentage for a particular response approaches 0% or 100%.

  Option 2 adds to Option 1 a requirement for the report to include a table, or narrative equivalent, indicating the margin of sampling error for other observed percentages (i.e. not just for an observed percentage of 50%) and for a representative selection of sample sizes. For example, this could be done as a single table, with sample sizes or key subgroups on one dimension, and a selection of observed percentages on the other dimension.

  The reason some Panelists preferred Option 1 over Option 2 is that Option 2 requires a bit more work. However, some Panelists did not feel strongly about this, and could live with either option.

The following options were each supported by one Panelist:

- *Option 3:*

  For probability samples, state the level of precision, including the margin of sampling error and confidence interval for the total sample for every question summarized in the survey report, and for any subgroup results reported for each question.

  The rationale was that with modern data processing software this information is readily available. Some Panelists, however, felt this amount of detail is excessive and not necessary.

- *Option 4:*

  For probability samples, state the level of precision, including the margin of sampling error and confidence interval for the total sample for an observed percentage of 50%.

  Results for subgroups will have a larger margin of sampling error because of their smaller sample sizes. Also, the margin of sampling error is highest for questions where 50% of the respondents gave one answer and the other 50% gave another answer. The margin of sampling error decreases as the observed percentage for a particular response approaches 0% or 100%.

  The rationale for this option is that it is simple, and is often how margin of sampling error is currently reported for surveys.

Statistical Treatment of Census Survey Results

Background and Questions: Possible consistency issue in Section 4.6 *Census Surveys*

PORD has received some industry feedback that part of what the current standards say about the statistical treatment of census survey results may not be correct.

The relevant language is in Section 4.6 *Census Surveys*. The following is an abridged version of 4.6:

> ### 4.6. Census Surveys
>
> In a census survey, an attempt is made to collect data from every member of a population. For example, an organization might want to do a survey of all of its employees. In this case, the population is "all of the organization's employees", and this would qualify as a census survey if all employees are invited to participate in the survey.
>
> The list whereby all members of the target population are to be contacted and invited to respond must be clearly described, including any of its limitations/exclusions in representing that target population. Whenever possible, an estimate of the percentage of the population that is excluded from the list must be provided and the potential impact of their exclusion on the research results considered. …
>
> 1) The number of attempted re-contacts and procedure for attempted re-contact must be stated.
>
> 2) Do not state a margin of sampling error, as this does not apply to census surveys because no sample is drawn.
>
>   The survey report must contain a statement on why no margin of sampling error is reported, based on the following template: "Since the entire population of [target population] was invited to participate in this study there is no margin of sampling error to be estimated or reported. The potential impact of non-sampling error due to non-response is discussed in the results section of the report. [If weighting was done, state the following sentence on weighting:] The data have been weighted to reflect the composition of [the target population (if known) or the sampling frame (e.g., client-supplied list)] on the main known characteristics."
>
> 3) There is no need to use inferential statistical tests since the results (frequencies, percentages) reported in a census survey describe the entire target population. — Statement #1
>
>   However, it is acceptable to use statistical significance tests to measure differences between subgroups within the target population. — Statement #2
>
>   As with any surveys, be they sample or census, the impact on the results of non- sampling error due to non-response must be assessed to the extent possible, and appropriate caveats on the interpretation of the results must be clearly stated.

The feedback was that Statement #2 is not consistent with Statement #1: *"Using the same reasoning, a subgroup of the census would still be a census of the subgroup. Statistical significance tests would only be appropriate on random samples of the census survey."*

Questions Addressed by the Panel

The Panel was asked whether the two statements flagged above are inconsistent.

Advisory Panel Response

Most Panelists said that Statement 1 is fine, but Statement 2 is not appropriate. The reason for the latter is that if a survey is a census or attempted census of a population, then it is also a census or attempted census of the subgroups within that population. Therefore inferential statistics tests are not needed for either the total sample or subgroups of the sample.

A few Panelists noted that there is a school of thought that statistical significance tests using census survey data may be legitimate when trying to determine if differences among groups are meaningful, but this is not a universally accepted practice. It was suggested that when using census survey data, a more useful way to determine if a difference is meaningful is to focus on the effect size and consider whether the effect size is meaningful given the policy or issue involved.

Background and Questions: Does response rate affect whether or not a survey is a "census"?

Questions Addressed by the Panel

In a census, an attempt is made to contact every individual in the population (subject to coverage error). However, rarely will an interview be completed with every attempted contact – that is, the response rate will usually be less than 100%.

The Panel was asked to consider whether or not a survey ceases being a census/attempted census if response rate falls below a certain level.

Advisory Panel Response

A census with a response rate of less than 100% is still a census, albeit one better described as an attempted census. Margin of sampling error does not apply. However, other sources of survey error can still be present. In particular, there can be non-response bias affecting survey measures if those not responding to the survey are systematically different from responders in some meaningful way.

Required Questions in Surveys

Section 2.1.3 of the Online and Telephone Standards gives required questions that must be asked in all surveys of individuals, *"unless a convincing argument is made that the research objectives are better served by excluding one or more of them"* (there is also an exclusion for business-to-business research where the unit of analysis is the organization).

Section 2.1.3 states:

> The wording used for each question must be that provided below, unless a convincing argument is made that particular research objectives require alternative wording. Even in these exceptional cases, the terms used and/or categories applied (e.g., for household income) to capture responses must be those provided below.

Section 2.1.3 states the following rationale for these required questions:

> The data from the age, education, and language questions (along with the recording of geographic location and sex) allows comparison with Statistics Canada census data* for the purpose of non-response analysis. The data, along with that from the employment status and income questions, also facilitate the comparison of results between Government of Canada public opinion research studies. (See Section 8. for further detail on non-response bias.)
>
> *A Panelist suggested revising this to "…allows comparison with official Statistics Canada population counts…" The reason is that between censuses, Statistics Canada updates population demographic counts using multiple sources of data. So, the latest population demographic counts may not based solely on the previous census.

Comparability to Statistics Canada's questions and response options is a critical requirement. This is needed for non-response analyses, and it is needed when the variables are used to weight survey data to match the population.

Because of differences between the online (self-completion) and telephone (interviewer-administered) modes, the questions/response options may be somewhat different for the two modes.

Required Questions: Introduction wording

Background and Questions

There is a general requirement (Section 2.1.2) to inform respondents at the beginning of a survey of the confidentiality of their questionnaire responses, but the current standards do not state any specific wording for how to preface the block of demographic questions located near the end of the questionnaire.

For reference, the relevant parts of Section 2.1.2 are:

> 2.1.2 The following are required elements of all Government of Canada online survey questionnaire introductions:

f) inform respondents that their participation in the study is voluntary and completely confidential;

g) inform respondents that their responses remain anonymous. In the exceptional cases where the research objectives require that respondent identity be revealed, the informed consent of the respondent must be obtained;

The Privacy Commissioner has requested the addition of a statement on privacy before the demographics section at the end of a questionnaire, such as this:

> *These last few questions are for statistical purposes and will be kept confidential. Your identity will always remain anonymous.*

Questions Addressed by the Panel

The Panel considered possible wordings for an introduction to the block of demographic questions at the end of a questionnaire. For discussion purposes, the Panel started with the following proposed text:

2.1.3 The following statement is required for all Government of Canada telephone/online surveys prior to administering the demographic section of the questionnaire:

> *These last few questions are for statistical purposes only. Your answers will be kept anonymous and confidential and will be combined with the answers from other respondents to this survey.*

Advisory Panel Response

With regard to the suggested introductory phrase, *These last few questions are for statistical purposes only*, the Panel preferred the following wording instead: *These last few questions will allow us to compare groups of respondents.* It was felt that *"compare groups of respondents"* would be more easily understandable by respondents than *"statistical purposes."*

With regard to the follow-up introductory sentence, the Panel was split across the following alternatives:

- Option 1: *Your answers will be kept anonymous and confidential.*
- Option 2: *Your answers will be kept anonymous.*
- Option 3: *Your answers will be kept confidential.*

Note that in all options, the proposed phrase, *and will be combined with the answers from other respondents to this survey*, is not present. It was felt that this adds to survey length without providing any much added meaningful information to respondents.

Choice of *"anonymous"* and/or *"confidential"* depends on two factors:

- What the intended meanings of these words would be if used in the Standards
- How respondents actually interpret these words when seen or heard in a survey questionnaire

How respondents would interpret these words in a survey context is not really known, and it is possible that there are individual differences in interpretation. It should also be kept in mind that in

the context of a survey questionnaire, a respondent may not spend much time analyzing the meaning.

While respondents' interpretations may not really be known, the GC Standards should be based on clear intended definitions when choosing the appropriate language. The following is an example of how one might think about these terms:

- "Anonymous" could mean a response is not linked to anything that personally identifies the respondent. From this perspective, "anonymous" may or may not apply to a data set. For example, a raw data file from a POR survey may contain personally identifying information (e.g. a phone number or email address), meaning that the responses in the data file are not anonymous. However, in a later data processing step, the personally identifying information may be removed from the data file, at which point the responses are anonymous.
- "Confidential" could mean that there are restrictions on who is able to access the data. For example, a reason some Panelists preferred using only the word "anonymous" (Option 2) focused on reporting of verbatim responses to open-ended questions. Selected verbatim responses might appear in a report, and survey data files provided to the GC client may contain all of the verbatim responses. While the respondents might be anonymous (i.e. not personally identified), some survey respondents may not consider this as keeping their contribution "confidential." This gets into the issue of how respondents interpret "confidential."

  A different interpretation of "confidential" is that it applies to responses that are linked to personally identifying information and it means that there are restrictions on who is able to access the data. To the extent that at some stage of data collection and processing POR survey data files contain personally identifying information, some Panelists preferred to use the word "confidential." In this context, one suggestion was to pair the word "confidential" with an explanation that any data shared with the GC or other third party will be anonymized.

The key point is that the decision about use of "confidential" or "anonymous" in the introduction to the block of demographic questions near the end of the survey will need to be based on clear intended definitions.

Use of *Prefer not to answer* in the Required Questions for Online Surveys

In the current required demographic questions for telephone and online surveys, there is a notable difference in the response option, *Prefer not to answer*, and related forms of this response option – i.e. *Don't know*, *Refused*:

- Telephone: The interviewer only reads the relevant demographic response options. The *Prefer not to answer* option (or its related forms) is never read to the respondent, but the interviewer is required to record this outcome if the respondent volunteers it.
- Online: *Prefer not to answer* is explicitly listed as a response option in all of the required demographic questions.

There was agreement by the Panel with the approach used on telephone surveys – that is, the interviewer does not read the *Prefer not to answer* response option (or its related forms), but does record this if volunteered by the respondent.

There was disagreement about whether *Prefer not to answer* should always be presented as an option in online surveys:

Do not require presenting the *Prefer not to answer* option in online surveys

Reasons for this position included:

- Respondents are more likely to answer by choosing one of the relevant demographic categories.
- Respondents can still behaviourally express a "prefer not to answer" response by skipping the question. At the researcher's discretion, the introduction to the demographic questions could include a statement telling the respondent they can skip a question, if the researcher is concerned respondents may not realize they have this choice and the researcher wants to give them this choice.
- Statistics Canada online surveys usually do not give a *Prefer not to answer* option, although the respondent may be allowed to skip the question.
- The mandated response options are the same for both telephone and online surveys, making it more straightforward to combine results in a multi-mode survey, or to compare data from different surveys using different modes.
- If a demographic question is being used near the beginning of the survey questionnaire to implement quota controls or as a filter determining subsequent questions, it is better to require a valid response. In this regard, one suggestion was to preface these quota/streaming control questions with something like the following to encourage responding: "*To ensure the people interviewed match the overall distribution of the general public, it is important to know the [insert demographic dimension] of all respondents*."
- If a required demographic question is used in weighting and there are respondents who *Prefer not to answer* (or otherwise do not answer), accommodations have to be made in order to apply weighting.

Explicitly state the *Prefer not to answer* option in online surveys

Reasons for this position included:

- Online GC POR surveys are usually programmed so that a respondent has to answer a question in order to be able to progress to the next question. This can create a conditioning effect, such that by the time a respondent reaches demographic questions near the end of the questionnaire, they have been "trained" to think a response is required even if in fact it is not required for these questions. If respondents think they are required to answer these questions, this can potentially have negative impacts on data quality and/or how the respondent feels about the survey. A respondent may not like feeling "forced" to answer a question they do not want to answer, and they may even deliberately give an inaccurate answer or drop out of the survey.

- If there is no *Prefer not to answer* option and a respondent doesn't answer the question, it can be unclear whether they deliberately decided not to answer or if they just accidentally hit the "Next" button and skipped the question.
- Usage of *Prefer not to answer* is often relatively high for the household income question. In most cases it is preferable to keep these respondents in the sample, so for this question it is particularly important to offer an explicit *Prefer not to answer* option.

The Panel concluded that *Prefer not to answer* (and its related forms) should be removed as a required listed response option in online surveys, but rather retained as an optional response category. In online surveys where there could be a conditioning effect such that respondent's may not realize they can choose not to answer, it would be a good practice to inform them of when they have this option.

Efficiency of Reading a Large Number of Response Options for Required Questions in Telephone Surveys

Some of the required demographic questions have a relatively long list of response options. This includes the questions on age, education and household income.

The Panel's view is that in a telephone interview, the interviewer should not be required to read all the response options before accepting a response, but rather should be allowed to tell the respondent to stop at the category that applies to them. This can make the telephone version of the question wording a bit different from the wording in the online version, but the benefit is a reduction in survey duration for telephone surveys.

Using the household income question as an example, the telephone version of the question could be revised to something like the following: *Please stop me at the category that best describes your total household income, that is, the total income of all persons in your household combined, before taxes last year. (READ LIST)*

The suggestion is to revise Section 2.1.3 to allow researchers the option of modifying the wording of a demographic question to allow the interviewer to instruct a respondent to stop at the category that applies to them. For example, the required wording of a question could be modified by prefacing it with an introduction of the form, *Please stop me at the category that best describes your [insert demographic dimension]*.

Required Questions: Gender

Background and Questions

The following is the current mandated question in Section 2.1.3 for telephone and online surveys:

Telephone Surveys	***Online Surveys***
Gender: [Do not ask: record based on interviewer observation]	**Gender**: What is your gender? Female Male Prefer not to answer

In the currently mandated questions, gender information is collected very differently in telephone versus online surveys. The telephone version does not actually ask a question about gender, but rather relies on interviewer observation. The online version asks a question and so is based on respondent self-classification rather than interviewer classification. Also, unlike the online version, the telephone version logically does not have a *"prefer not to answer"* response option (the research firm might have an "indeterminate" response option, though it is not known to what extent firms actually give interviewers this type of option).

Considerations around phrasing of this question include:.

- Statistics Canada's 2016 census used the following question – although as noted below this will change in the next census:

  What is this person's sex?
  Male
  Female

  The census questionnaire in other major countries (e.g. U.S., United Kingdom, Australia) also asked a similar question and offered only the two answer options shown above on their most recent census data collection cycle.
- There is a difference between a person's sex and a person's gender. While they are related, they are not the same thing. Sex is binary and refers to one's biology at birth whereas gender is not binary and can often be about how one identifies oneself physically and/or psychologically. The current mandated online survey question asks about "gender", while the 2016 census question asked about "sex".
- Even if a question asks for the respondent's "sex", they may answer based on gender – and there is no way knowing which approach they took.
- The census bureaus in all the major countries noted above, including Statistics Canada, have committed to including a question(s) to accurately capture 'gender' identity for their next census, and all have started research to test questions in preparation for their next census.
- There are various organizations (e.g. Equality and Human Rights Commission, Rainbow Health Ontario, the Williams Institute) and a private sector firm – GfK Research – that have looked at how best to ask these questions, but there is no general consensus on either how best to ask the questions or what answer options should be provided to accurately capture the information.

- In January 2016, the U.K. market research standards organization, Market Research Society (MRS), published *MRS Guidance on Collecting Data on Sex and Gender*. In their guidance they state:

  > It is important that all research participants feel satisfied that their voice is being heard. Allowing participants to provide information in a way that reflects the view they want to express is explicitly set out in the MRS Code of Conduct and this also applies to describing their sex or gender. The categories provided for completion must facilitate and recognise the fluidity of gender identity, by providing an additional option such as free-field 'other'.
  >
  > Members must ensure that participants will be able to proceed through any research without being required to indicate male or female options if they do not identify as such. This may be achieved by including female, male and other as (free field) options; allowing participants to self-identify by asking an open-ended question or by adopting a two-step approach that covers assigned sex at birth and then current gender identity.

  MRS also notes the following key points:

  - ✓ Collection of sex and/or gender information needs to be tailored to the data collection tool.
  - ✓ Self-completion tools should at a minimum include an additional field of "other" for either gender or sex questions.
  - ✓ Interviewer-aided techniques for collecting information need to build in a level of discretion and flexibility for interviewers, to ensure questions are responsive and are handled appropriately.

  MRS provided the following examples of questions and a checklist of what questions researchers should ask themselves before deciding what question to ask:

What is your sex?	*What is your sex? OR What is your gender?*
Male *Female* *Intersex* *Prefer not to say*	*Male* *Female* *Other (please specify)* *Prefer not to say*

  Checklist - Questions to ask

  Do I need to collect information on either sex or gender? Is there a research purpose for collecting this information?
  - o Why?
  - o Is it relevant and not excessive?

  Should I collect information on sex or should I collect gender?
  - o Do I want to match with other data sets or sources? If so, what do they collect?

  Can I make the answers optional?

  What response options should I provide?
  - o Should I provide closed categories (including other) or an open field?
  - o Do I want to include an intersex option for sex or an "other" option?

  The MRS goes on to say the following about gender identity questions specifically:

  > Establishing best practice in developing and asking gender identity questions will need to build on the position and practice of the ONS [Office of National Statistics] and research carried out by the EHRC [Equality and Human Rights Commission]. This will allow the research community to design and implement a consistent and standard gender identity

question that can be understood and answered by all people living in the United Kingdom (UK).

- Uptal Dholakia in his article *How Should Market Researchers Ask About Gender in Surveys*? in an online blog for Psychology Today (September 2016) points out that:

> Any well-designed market research survey is based on two core principles: the ***principle of accuracy*** and the ***principle of inclusiveness***. A questionnaire should be designed to gather information accurately, using best practices of survey design that psychometricians have formulated over several decades. But this is not enough. A survey should also be inclusive. When a respondent has finished taking a survey, they should feel like the opinion they have provided will be valued just as much as every other survey taker.

Questions Addressed by the Panel

The Panel considered whether or not the current required gender questions should be revised, and what revisions might be appropriate.

Advisory Panel Response

The mandated question should ask for the respondent's gender, not sex, and provide the response options of *Male*, *Female*, and *Other*.

Asking for gender rather than sex respects the principle of inclusivity, and is consistent with Statistics Canada's plan to ask for gender, with an "other" response option, in the next census.

Until the next census data are available, there will be a mismatch between this revised gender question on GC POR surveys and the existing census data based on a question about sex. However, according to a Panelist, Statistics Canada has found relatively little difference in responses to the two types of question, and relatively low usage of the "other" response option. Therefore, the impacts on weighting are minor and can be managed.

Because of the addition of the *Other* response option, the gender question must be asked explicitly in telephone surveys, rather than left to interviewer observation. Asking the question explicitly also respects the participant's self-classification of their gender. The result is that the same question would be asked in both telephone and online surveys:

Telephone Surveys and Online Surveys
Gender: What is your gender? Female Male Other

A few Panelists who supported asking this question in both telephone and online surveys commented that in telephone surveys this question may lead to some complaints, jokes or refusals, and interviewers will need to be prepared for this. For example, a suggestion was to allow an interviewer to "soften" the question by preceding it with something like *"just to confirm…"* This

type of “softening” preserves the required question wording and response options, and therefore would be consistent with the suggested required wording.

Another possible approach to “softening” this question in a telephone survey would be to not require reading the response options. That would mean the “other” category is selected by the interviewer whenever a respondent gives an answer that is neither female only nor male only. While this results in a different presentation of the question in telephone versus online surveys, it seems reasonable to expect that the results would be comparable – i.e. that this would not cause a mode effect.

It was noted that depending on the objectives of some surveys, the *Other* response option might be expanded into an open-ended *Other – please specify* option, or expanded into a longer list of response options. This optional wording would still be consistent with the basic requirement to have an *Other* response option.

One Panelist observed that respondent reaction to the use of the English word *gender* may be different from reaction to the French word *genre* in *Quel est votre genre?* The thought was that this may be perceived as more awkward or odd in French. Statistics Canada has indicated that they have moved to using the word *genre* in some recent surveys, albeit with (a) an explanation and (b) preceded by a question on the individual’s sex:

> Q1 : Quel était votre sexe à la naissance? Par sexe on entend le sexe assigné à la naissance.
>
> Q2 : Quel est votre genre ? Par genre, on entend le genre actuel, qui peut différer du sexe assigné à la naissance ou de celui inscrit dans les documents légaux.

Assuming *genre* is used in the next census, then for comparability the same wording should be used in the Standard for the French language version of the gender question.

One Panelist had a different opinion on what should be the mandated question, and suggested using a multi-step question that first establishes sex at birth, and then establishes gender. This would look something like the following:

1) What was your sex at birth?
   Female
   Male
2) Do you currently consider yourself to be [insert Q1 answer]?
   Yes
   No
3) [If “no” at Q2, ask:] Do you consider yourself to be (read list):
   The list of options could be customized to the survey.

Required Questions: Language

Background and Questions

The following are the two mandated questions for language for telephone and online surveys in Section 2.1.3. The researcher can choose to use both questions or only one of the two questions, depending on the survey objectives:

Telephone Surveys	***Online Surveys***
Mother Tongue: What is the language you first learned at home as a child and still understand? [READ LIST — ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] [DO NOT READ] Don't know/Refused	**Mother Tongue:** What is the language you first learned at home as a child and still understand? [ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] Prefer not to answer
Language Spoken at Home: What language do you speak most often at home? [READ LIST — ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] [DO NOT READ] Don't know/Refused	**Language Spoken at Home:** What language do you speak most often at home? [ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] Prefer not to answer

The Panel was asked to provide input on what question or questions to require in GC POR surveys for official language of respondents.

Considerations include:

- There is a need to allow for comparison of data collected on language to a high quality data source, e.g. Statistics Canada census data, for the purpose of non-response bias analysis or weighting. To perform these types of analyses, the questions should be asked in the same way in POR surveys as they are in the high quality data source.
- The 2016 census asked the following language questions:

  Q.7: *Can this person speak English or French well enough to conduct a conversation?*
  English only
  French only
  Both English and French
  Neither English or French

  Q.8a: *What language does this person speak **most often** at home?*
  English
  French
  Other language -- specify

Q.9: *What is the language that this person **first learned** at home **in childhood** and **still understands**?*

English
French
Other language -- specify

- *Language spoken most often at home* (Q.8a) is more useful as a fundamental measure of language use than is Q.7, so it is not suggested to make a question like Q.7 mandatory in surveys.
- PORD obtained some input on this matter from the MRIA. The general view expressed in that input was that the question *language spoken most often at home* is considered to be the more useful indicator of official language, and mother tongue was seen often to be less relevant. This is not to say that asking both questions would never have value, but rather is a point of view on which of the two existing questions is more often likely to be useful.

Questions Considered by the Panel

The Panel was asked to consider whether both of the existing required language questions should be retained, and to comment on wording and response options. Note that the discussion below does not deal with the issue of whether or not to include a *Prefer not to answer* option, as this issue was discussed earlier.

Advisory Panel Response

The Panel felt that *Language used most often at home* is the most useful form of the language question and should be the only required question. Adding *Mother tongue* would be at the discretion of the researcher. The current wording of the *Language used most often at home* question is appropriate.

Required Questions: Age

Background and Questions

The current required age question is:

Telephone Surveys	***Online Surveys***
In what year were you born? [Record year – XXXX]	In what year were you born? [YYYY] Prefer not to answer
[IF PREFERS NOT TO PROVIDE A PRECISE BIRTH YEAR, ASK:] Would you be willing to tell me in which of the following age categories you belong?	[IF PREFERS NOT TO PROVIDE A PRECISE BIRTH YEAR, ASK:] Would you be willing to indicate in which of the following age categories you belong?
18 to 34 35 to 49 50 to 54 55 to 64	18 to 34 35 to 49 50 to 54 55 to 64

OR 65 or older? [DO NOT READ] Refused	65 or older Prefer not to answer

A survey can use more detailed breaks as long as these can be collapsed into the categories above.

In the census, Statistics Canada determines age from birth date.

For reference, the 2016 Census Profile for Canada shows the following age distribution:

Table 1 Statistics Canada Census Profile (2016): Age Distribution

Age Range	% of Population 18+		
18-24	10.9%	18-24	10.9%
25-29	8.1%	25-34	16.4%
30-34	8.3%		
35-39	8.1%	35-44	16.2%
40-44	8.0%		
45-49	8.4%	45-54	17.9%
50-54	9.5%		
55-59	9.3%	55-64	17.5%
60-64	8.1%		
65-69	7.0%	65 or over	21.1%
70-74	5.1%		
75-79	3.6%		
80 or older	5.4%		

Questions Considered by the Panel

The Panel was asked to consider whether there should be any changes to the required age question, including possible revisions to the break points or whether a smaller number of age categories should be used.

Advisory Panel Response

The mandated age question is structured appropriately in that it first asks for date of birth, and only presents age categories if the respondent declines to give their date of birth. Most respondents will answer the date of birth question.

With regard to the follow-up question addressed to respondents who decline to give date of birth, there were suggested changes to the categories:

- The current two age breaks for the 35-54 age range are 35-49 and 50-54. This should be changed to 35-44 and 45-54 in order to get a more even distribution of age categories.
- Some Panelists disagreed with using an 18-34 age range, and recommended this be split into 18-24 and 25-34. Their argument was that the 18-24 age range can be less likely to

participate in surveys, and so are at risk of being under-represented. Breaking out the 18-24 age range makes it possible to detect and correct for under-representation of this age group.

Required Questions: Education

Background and Questions

The current required education question is:

Telephone Surveys	*Online Surveys*
What is the highest level of formal education that you have completed? [READ LIST]	What is the highest level of formal education that you have completed?
Grade 8 or less	Grade 8 or less
Some high school	Some high school
High School diploma or equivalent	High School diploma or equivalent
Registered Apprenticeship or other trades certificate or diploma	Registered Apprenticeship or other trades certificate or diploma
College, CEGEP or other non-university certificate or diploma	College, CEGEP or other non-university certificate or diploma
University certificate or diploma below bachelor's level	University certificate or diploma below bachelor's level
Bachelor's degree	Bachelor's degree
Post graduate degree above bachelor's level	Post graduate degree above bachelor's level
[DO NOT READ] Prefer not to answer	Prefer not to answer

For reference, the 2016 Census Profile for Canada shows the following education categories:

Table 2 Statistics Canada Census Profile (2016): Education

No certificate; diploma or degree	18.3%
Secondary (high) school diploma or equivalency certificate	26.5%
Postsecondary certificate; diploma or degree	55.3%
Apprenticeship or trades certificate or diploma	9.8%
College; CEGEP or other non-university certificate or diploma	19.4%
University certificate or diploma below bachelor level	2.8%
University certificate; diploma or degree at bachelor level or above	23.3%
Bachelor's degree	15.5%
University certificate or diploma above bachelor level	1.6%
Degree in medicine; dentistry; veterinary medicine or optometry	0.7%
Master's degree	4.6%
Earned doctorate	0.8%

It has been pointed out that the first two categories in the current required education question do not line up well with the Quebec education system. PORD provided the following from Wikipedia:

> *Mandatory elementary education (école primaire) starts with grade 1, through to grade 6. Secondary school (école secondaire) has five grades, called secondary I-V (Sec I-V for short) or simply grades 7-11. Students are 12 to 16 years old (age of September 30), unless they repeat a grade. Upon completion of grade 11, students receive their high school diploma from the provincial government.*

In Quebec, Grade 8 is the start of high school (école secondaire), whereas outside Quebec Grade 9 is the start of high school. The view is that the first two existing required categories – *Grade 8 or less*, and *Some high school* – are confusing in Quebec, and indeed the two categories overlap, since Grade 8 is also "some high school" in Quebec.

A possible revised approach is to drop any attempt to distinguish subcategories of less than a high school diploma (such distinctions can always be added on an ad hoc basis for a particular survey). In this scheme, the first two response options could be:

Less than a high school diploma or equivalent
High School diploma or equivalent

Questions Addressed by the Panel

The Panel considered whether and how to modify the education question, including the proposed revision to combine the two "less than high school" response options into one category.

Advisory Panel Response

The Panel supported combining the two "less than high school" response options into a single category, *Less than a high school diploma or equivalent*. It was noted that these are usually combined when examining the relationship of education to other survey variables. A researcher would still have the option of breaking the "less than high school" category into subcategories should that be useful for a particular survey.

One Panelist commented that it is not uncommon for young adults to start but not finish postsecondary education, or to start postsecondary and then take a break for a period of time, or to attend more than one postsecondary institution over a period of time before getting a degree. Depending on the point in time they are asked for education, they have more than a high school education but do not have a degree or certificate. For some purposes, it can be useful to include this as a response option. Note that this is still consistent with the mandated education question, in that this subcategory response option can be recoded into the mandated response options.

Required Questions: Household Income

Background and Questions

The current required household income question is:

Telephone Surveys	*Online Surveys*
Which of the following categories best describes your total household income? That is, the total income of all persons in your household combined, before taxes [READ LIST]? Under $20,000 $20,000 to just under $40,000 $40,000 to just under $60,000 $60,000 to just under $80,000 $80,000 to just under $100,000 $100,000 to just under $150,000 $150,000 and above [DO NOT READ] Refused	Which of the following categories best describes your total household income? That is, the total income of all persons in your household combined, before taxes? Under $20,000 $20,000 to just under $40,000 $40,000 to just under $60,000 $60,000 to just under $80,000 $80,000 to just under $100,000 $100,000 to just under $150,000 $150,000 and above Prefer not to answer

It is assumed a survey could optionally use more detailed breaks as long as these can be collapsed into the categories above.

Statistics Canada now sources income data from administrative records, although some Statistics Canada surveys do ask some form of income question.

Statistics Canada reports various income measures, one of which is “total income for private households before tax”, which appears to be equivalent to “total household income” in the questions above.

For reference, based on 2015 data, the Census Profile for Canada shows the following for total income for private households:

Table 3 Statistics Canada Census Profile (2015): Total Income of Private Households

Census Profile for Canada		*Current Required Breaks*	
Under $5;000	1.6%	Under $20,000	9.7%
$5;000 to $9;999	1.4%		
$10;000 to $14;999	2.7%		
$15;000 to $19;999	4.0%		
$20;000 to $24;999	4.3%	$20,000 to just under $40,000	16.7%
$25;000 to $29;999	3.8%		
$30;000 to $34;999	4.3%		
$35;000 to $39;999	4.3%		
$40;000 to $44;999	4.2%		16.1%

$45;000 to $49;999	4.1%	$40,000 to just under $60,000	
$50;000 to $59;999	7.8%		
$60;000 to $69;999	7.2%	$60,000 to just under $80,000	13.7%
$70;000 to $79;999	6.6%		
$80;000 to $89;999	5.9%	$80,000 to just under $100,000	11.2%
$90;000 to $99;999	5.3%		
$100;000 to $124;999	10.4%	$100,000 to just under $150,000	17.7%
$125;000 to $149;999	7.2%		
$150;000 to $199;999	7.9%	$150,000 and above	14.7%
$200;000 and over	6.8%		

Some other statistics using 2015 data:

- Median total income for private HH: $70,336
- Average total income for private HH: $92,764
- Percentage of Canadians living in low-income (LIM-AT[8]): 14.2%

  The dollar threshold for low income varies with household size. Statistics Canada has the following table showing the low-income thresholds as a function of household size[9]:

Table 4 Statistics Canada Census Profile (2015): Low Income Measures Thresholds

HH Size	After-tax income	Before-tax income
1 person	$22,133	$25,516
2 persons	$31,301	$36,084
3 persons	$38,335	$44,194
4 persons	$44,266	$51,031
5 persons	$49,491	$57,054
6 persons	$54,215	$62,500
7 persons	$58,558	$67,508

Questions Addressed by the Panel

The Panel considered whether there should be any revisions to the household income question or response options. Note that presentation of a *Prefer not to answer* response option is discussed elsewhere in the report.

[8] LIM-AT is a low-income measure based on after-tax household income

[9] Statistics Canada, Table 4.2, Low-income measures thresholds (LIM-AT and LIM-BT) for private households of Canada, 2015 http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/tab/t4_2-eng.cfm

Advisory Panel Response

The Panel recommended that the question wording be revised to specify a time frame. This reduces ambiguity and uncertainty in how to answer the question among respondents whose household income has changed recently. The specific suggestion was to use a "last year" time frame.

A few Panel members also suggested that a reminder to respondents that household income can come from a variety of sources could provide more accurate data. One Panelist suggested a detailed list of different income sources be included and another Panel member suggested that the reference be more generic, i.e. total income from all sources for all people in the household.

The following is a possible question that incorporates both the time frame and the "all income sources" cue: *Which of the following best describes your total household income before taxes last year from all sources for all household members?*

The Panel agreed with retaining the existing response options.

Possible Additional Required Question for Telephone Surveys: Type of Phone(s) in the Household

Background and Questions

PORD consulted with MRIA on the value of adding a question in telephone surveys about the type(s) of phone(s) a respondent has access to at home:

> **PORD question to MRIA**:
>
> Should we consider adding questions on landline vs. cell for classification purposes (landline only; cell only; both) e.g. adding a question at the beginning of the survey similar to the question to cell phone respondents: *At home, do you have a cell phone as well as a traditional telephone line?*
>
> **MRIA comments and response**:
>
> This question is valid in that it may help the researcher to know if there is a segment of the population in question that was missed. The ratio of landline contacted to cell phone contacted based on the known demographics of landlines and cell phone lines in the area in question may provide additional insights in the analysis.

The information about phone usage may also be incorporated into the weighting scheme for telephone surveys. The literature review commissioned by PORD notes this possibility (see pp. 13-14 of the review). The review says that, *"At the time of writing, there is no consensus on the best approach to weighting dual-frame survey samples."* It goes on to cite an example of an approach that weights by telephone status (cell phone only, landline only, dual phone users), and an approach that does not.

Classification of household telephone status can be quite complicated, depending on the approach taken to weighting, or to how respondents are selected for the interview within a household[10]. Some examples of complicating factors:

- People in a household with a landline may nonetheless most often use a cell phone ("cell phone mostly"), and similarly someone with a cell phone may mostly use a landline ("landline mostly").
- An individual may have more than one cell phone.
- People within a household may share a cell phone.
- A cell phone might be personal use only, business use only, or a combination of personal and business use.

Questions Addressed by the Panel

The Panel focused on whether and how to ask questions that establish up to a three-way classification of a respondent's household: landline only household, cell phone only household, and dual landline and cell phone household.

Advisory Panel Response

The Panel agreed with adding a required question on household phone status, but with the understanding that whether and how it might be used for weighting or quota controls will be decided for each POR project.

The following factors played a role in the Panel discussion of whether and how to determine household telephone status.

Whether or not household phone status should be used for weighting: The literature review commissioned by PORD noted that in 2018 there is no consensus on whether or how to use household phone status in weighting survey data, and referenced an approach that uses this type of weighting and an approach that does not. One Panelist noted that depending on the survey topic, household phone status may have no direct relationship to the survey variables of interest. For example, if respondent age (which is correlated with household phone status) has a more direct bearing on survey responses, then it is better to use age in weighting than an "indirect" correlate such as household phone status.

Another complicating factor is that the use of household phone status in weighting or quota control can be difficult because of a scarcity of high quality, up to date population information on household phone status in Canada. Up to date data are important because the relative usage of cell phones and landlines has been changing and continues to change.

[10] A good overview of the complications of weighting telephone surveys by telephone status can be found in 2010 AAPOR Cell Phone Task Force report, *New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers* – see the section on Weighting. The report also has an appendix showing the questions some major survey organizations were using at the time to ascertain telephone status. The report concluded that there was no consensus on what questions to ask to ascertain telephone status.

The Panel agreed with adding a required question on household phone status, but with the understanding that whether and how it might be used for weighting or quota controls will be decided for each POR project.

Use of sample information: There are sample frames for cell phones and landlines, so there is information in the sample about whether a number being dialed is a cell phone or a landline.

This information is not, however, sufficient to classify the respondent's household telephone status, as it only addresses the status of one particular phone number. There can be more than one phone number associated with a household – e.g. multiple cell phone numbers, and/or a combination of cell phone numbers and a landline number. Therefore, it is necessary to ask questions in the survey to determine the respondent's household telephone status.

If it is assumed that the sample information on phone type is correct (but see below), then it is possible to ask just one question to determine household phone status:

> **"One-question" approach based on using sample information**
>
> If from the cell phone sample, then ask if there is a landline phone in the household
>
> ***Or***
>
> If from the landline sample, then ask about cell phones in the household

The advantage of this use of sample information is that household telephone status can be determined with only one survey question, which is important for reducing survey duration. One Panelist favoured this approach because of the importance of keeping telephone surveys as short as possible.

While sample information about type of phone is good, it is not perfect:

- *Porting:* A number for one type of phone can be ported to another type of phone. When this occurs, it is mostly porting a landline number to a cell phone. Porting from a cell phone to a landline is rare, but is possible.
- *Future accuracy of cell phone and landline sample frames:* One Panelist suggested that as the number of cell phones in use increases and the number of landlines decreases, the landline phone exchanges may increasingly be used for cell phones.

The majority of Panelists opted not to use sample information about phone type in determining household phone status. In practical terms, this means that at least two questions need to be asked to determine household phone status.

Three-way versus two-way classification of household phone status: In the Panel discussion, there were references to usage of two different classification approaches:

> Two-way classification: Cell phone only; Not cell phone only
>
> Three-way classification: Cell phone only, Landline only, Dual cell phone and landline

Both require asking at least two questions[11] (assuming sample information is not used). Questions that produce the three-way classification will also produce the two-way classification. However,

[11] It is possible to determine status using a single question, but this requires a more complex question with more response options, and in the end may not really save much time. The Panel preferred asking simpler questions, which is the reason for stating at least two questions are required. Another problem with a single, more complex question is that it cannot also measure number of cell phones in the household.

questions designed to produce the two-way classification may not support the three-way classification. The Panel opted for questions that produce the three-way classification, with the understanding that if a researcher chooses to use the telephone status question for weighting or quota control, it is their choice as to which classification scheme to use.

"At least one cell phone in the household" versus "Number of cell phones in the household": To produce the three-way classification with respect to cell phones, all that is necessary is to determine if there is at least one cell phone in the household. However, some Panelists said that the "number of cell phones in the household" can have value. This approach answers the question of whether there is at least one cell phone in the household (zero cell phones versus one or more cell phones), but also determines the number of cell phones. Knowing the number of cell phones can allow a more accurate estimate of the probability that a household could be randomly selected from the cell phone sample frame. The Panel agreed on using the "number of cell phones in the household" measure.

How to refer to a landline phone: Historically, "landline" has been used to refer to the phone(s) connected to a telephone line. However, with the increasing incidence of cell phone only households, particularly among younger people, there's a question of whether one can assume everyone understands what "landline" refers to. Further, there is increasing use of internet-based phone service (VOIP service), and it is unclear whether people would consider these to be "landline" phones. The Panel recommends that the English questions on household phone status use "home phone" rather than "landline":

- The term "home phone" is now used by the major telephone and cable companies.
- "Home phone" probably more clearly encompasses both "telephone line" service and VOIP service.
- A researcher can optionally include a definition in the survey. An example given of a definition was: "Home phone service is different from using your cell phone as your home phone. This is a service where you pay a monthly fee to have a dedicated line and phone number for your home."
- It was suggested the French terminology could be "une ligne résidentielle" or "un service de téléphonie résidentielle."

"Landline" versus "home phone" in the Standards: The preceding discussion pertained to the terminology used in the required question on household phone status. The Panel was also asked whether standards that refer to "landline" should continue to do so, or switch to "home phone" (e.g. in the current Telephone Standards, see 1.2.4 #7, 4.1.1, and 15.2). The Panel was split on this issue, with some preferring to retain "landline" as being technically more precise, and others preferring "home phone" as being consistent with the recommend use of "home phone" in the required question on household phone status. If "home phone" is adopted, the Standards should include a definition of what the term means.

Question location: There needs to be flexibility in where the household phone status question, and all of the other required questions, are placed in the questionnaire. Depending on the needs of the survey, they may be placed in the final section of the questionnaire, or they may be placed earlier in the questionnaire if needed for quota control or filtering purposes. The Panel's understanding of Section 2.1.3 of the Standards is that this flexibility currently exists.

To summarize, the Panel suggested the following additions to telephone surveys for type of phone(s) in the household:

- The Panel agreed with adding a required question on household phone status.
- The Panel recommends the following questions for determining household telephone status:
  1) Does your household subscribe to a home phone service? (Yes, No)
  2) How many working cell phones does your household have? (0, 1, 2, etc.)

Use of Mobile Devices in Online Surveys

In an online survey, it is likely that a sizable percentage of respondents will use, or attempt to use, a mobile device (smartphone or tablet) to complete the questionnaire.

A 2014 AAPOR report on mobile technologies concluded:

> **Recognize If You Are Conducting Online Surveys, You Are Conducting Mobile Surveys:** A non-ignorable and growing percentage of respondents are now accessing online surveys via their mobile browsers (with estimates ranging from 8 - 23% depending on the study), resulting in higher abandonment rates and potentially greater measurement error among these mobile respondents.[12]

With the growth of smartphone ownership and use, the percentages of people completing online surveys on a mobile device has very likely increased since 2014.

In a summary of results from the 2016 General Social Survey (GSS), it was reported by Statistics Canada in *The Daily*[13] that three-quarters of Canadians 15+ own a smartphone, although there is substantial variation by age. Most 15-34 year-olds (94%) reported owning a smartphone, compared with 69% of those aged 55 to 64 and 18% of Canadians 75 years and older.

In the 2015 Sage Research report, *Best Practices for Improving Cooperation for Online Surveys*[14], the following are some of the conclusions drawn from a review of the research literature:

- Dropout rates among people using a mobile device to do online surveys are higher than among those using a computer.
- Designing the questionnaire to be mobile-friendly can substantially improve the response rate among smartphone users.
- Another potential contributor to higher dropout rates among people using a smartphone is that it usually takes longer to do a survey on a smartphone than on a computer.
- An issue is whether there are mode effects – e.g. does a mobile-friendly version of a question answered on a smartphone yield the same response distribution as computer-friendly version answered on a computer?

Some survey companies are implementing technologies to accommodate use of mobile devices in online surveys, for example:

- Detection of type of device and browser
- Templates for mobile-friendly question design for various types of questions
- Automatic adaptation of question design to the device, including adapting to screen size

---

[12] AAPOR. (2014). *Mobile technologies for conducting, augmenting and potentially replacing surveys: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research.* Deerfield, IL: The American Association for Public Opinion Research.

[13] Statistics Canada. (2017). "Life in the fast lane: How are Canadians managing?, 2016". *The Daily*, Tuesday, November 14, 2017.

[14] Sage Research Corporation. (2015). *Best Practices for Improving Cooperation for Online Surveys*. Prepared for the Public Opinion Research Directorate, Government of Canada. See pages 53-56.

The key point is that adapting questionnaire design for those using a mobile device can potentially improve the survey data in terms of coverage, response rate, reduction of non-response bias, and answer quality.

An online survey can take different approaches to the possibility mobile devices will be used[15]:

1) *Do not adapt the survey to mobile.* This means people attempting to do the survey on a smartphone will not be using a mobile-friendly version of the questionnaire, but rather a version designed for completion on the larger screen of a computer. This will likely produce higher drop-out rates among people using smartphones compared to when a mobile-friendly version of the questionnaire is provided.
2) *Block mobile device users from doing the survey on their device, and encourage them to complete the survey on a computer.* The downside to this approach is that substantial non-response could occur if many do not make the effort to switch to using a computer.
3) *Optimize the survey to be correctly displayed on the most common smartphones in use among the survey target group.*
4) *Have the survey be fully compatible to be taken on any device* – which requires a survey platform that can handle adapting the questionnaire to the full range of devices.

The current standards do not address the possibility and implications of an online survey being completed on a mobile device. The objective for the Panel was to provide input to revising the Standards to address these matters.

The matters considered by the Panel pertaining to usage of mobile devices to complete online surveys were:

- *1. Proposal Documentation:* Information that must be in the proposal about how usage of mobile devices will be handled in an online survey
- *2. Questionnaire Design:* Requirements and guidelines for a mobile-friendly online survey
- *3. Pre-testing:* Pre-testing the effectiveness of the mobile-friendly versions of the questionnaire
- *7. Data Collection:* Collecting data on type of device
- *14.6. Quality controls:* Look for differences in response distribution if questions are displayed differently on different devices

[15] Callegaro, M. (2010). "Do you know which device your respondent has used to take your online survey?". *Survey Practice*, 3(6).

Proposal Documentation Relating to Use of Mobile Devices in Online Surveys

Background and Questions

Questions Addressed by the Panel

The Panel considered possible revisions to the following two sections of *Proposal Documentation* in the Online Standards, neither of which currently refer to potential use of mobile devices:

- 1.2.2 *Sample/Sampling Details*
- 1.2.5 *Questionnaire Design*

Advisory Panel Response

Revision to Section 1.2.2 *Sample/Sampling Details* in Online Standards

The Panel said the default expectation should be that an online POR survey sample will include respondents using either a computer or a mobile device.

Because of the belief that this is now an industry norm, a few questioned whether this needs to be stated at all in the Standards. However, to be clear that this is an expectation, it is better to state it explicitly. One Panelist also commented that there can be some surveys which need to be restricted to computers, i.e. larger screens. The example given was a business survey which has accounting figures in columns. So, there can be cases where a survey sample needs to be restricted to certain device types.

To make this expectation explicit, as well as to give a brief rationale for the expectation, the recommendation was to add the following to Section 1.2.2 *Sample/Sampling Details* of the Online Standards:

> Respondents in an online survey must be able to complete the questionnaire on either a computer (desktop or laptop), tablet or smartphone. If the intent is to limit the sample to only some types of devices, describe the reasons for this since restricting device usage can potentially impact coverage of the survey's target population.

Revision to Section 1.2.5 *Questionnaire Design* in Online Standards

Given the default expectation that the sample for an online survey will include users of computers and mobile devices, the Panel said the default expectation should be that surveys have a mobile-friendly version of the questionnaire.

The suggestion was to include the following sort of text in Section 1.2.5 *Questionnaire Design* of the Online Standards:

> An online survey must have a mobile-friendly version of the questionnaire. If the intent is not to have a mobile-friendly version of the questionnaire, describe the reasons for this.

Mobile-Friendly Online Surveys and *Questionnaire Design*

There were three revisions/additions to Section 2 *Questionnaire Design* of the Standards considered with respect to questionnaire design in online surveys where mobile devices may be used:

1. Should there be a standard encouraging use of a common question design/layout across devices?
2. Should there be a different survey duration standard for mobile-friendly surveys?
3. Should there be guidelines on features of a mobile-friendly questionnaire, and if so then what guidelines?

Background and Questions: Should There Be a Standard Encouraging Use of a Common Question Design/Layout Across Devices

The 2015 Sage Research report, *Best Practices for Improving Cooperation for Online Surveys*, reviewed the research literature and concluded:

- Logically, using a different question design/layout on different devices introduces the possibility of a "mode effect" – i.e. a mobile-friendly version of a question answered on a smartphone might yield a different response distribution than a computer-friendly version answered on the larger screen of a computer.
- The research was inconclusive, basically because there hadn't been much research on the topic. It's also the case that it is hard to do conclusive research on this topic, because of the diversity of question types and the diversity of ways in which these can be rendered: research might be conclusive for a particular question type rendered in a particular way on mobile versus computer, but this may not generalize to other question types or to other design approaches.
- That said, there were several studies suggesting that there may not be substantial mode effects[16].

There are two approaches to integrating mobile-friendly question designs into an online survey:

- *Mobile first:* In this approach, the questionnaire is designed from the outset to be mobile-friendly, and the mobile-friendly design is used on both mobile devices and computers. The assumption is that a mobile-friendly design will work just as well on the larger screen of a computer[17]. Note that a mobile-first approach could – and should – use software to adapt the question display to different types/sizes of mobile devices.

[16] For example see the following, which include both research reviews and experimental studies:

Toepoel, V., & Lugit, P. (2014). What happens if you offer a mobile option to your web panel? Evidence from a probability-based panel of internet users. *Social Science Computer Review*, *32*(4), pp. 544-560.

Wells, T., Bailey, J., & Link, M. (2014). Comparison of smartphone and online computer survey administration. *Social Science Computer Review, 32*(2), pp. 238-255.

[17] For example:

Saunders, T. (2015). Improving the Survey Experience for Mobile Respondents. *Alert! Magazine*, Third Quarter 2015, Marketing Research Association

Thomas, R. & Barlas, F. (2018). *Best Practices in Mobile-First Design*, GfK Whitepaper

- *Responsive design:* Each question design is optimized for mobile and optimized for computer, which means that the question design can be different for mobile (small screen) versus computer (large screen). The idea is that optimizing for each device results in getting the best quality data from each. This approach was recommended in a 2017 research article by Antoun, Couper & Conrad[18]:

  > These results have practical implications for researchers conducting Web surveys. Our finding of near-comparability in the quality of responses between smartphone and PC users, even for sensitive and burdensome questions, suggest not only that smartphone users should be accommodated in Web surveys but also that the survey instrument should be optimized for small screens. There is a line of thinking, based on *universal mode* design or *unimode* design principles (Dillman 2000), that one should aim to display the exact same question formats across devices in order to minimize mode effects. Our results seem to suggest otherwise since we found no evidence that the redesigned (optimized) question formats in the smartphone version introduced device effects or had an adverse effect on data quality.
  >
  > The notion of comparability that we would instead argue for is a *best practices* approach … This places the emphasis not on presenting the identical surface-level features, but on using the best practices of each mode or device in order to present the same stimulus to the respondent and minimize error within each mode or device. Carrying out these principles for mobile Web research requires playing to the strengths of smartphones by identifying and using input tools that are user friendly and avoiding any formats that are not (e.g., sliders, drop boxes that turn into pickers). This may also entail using Responsive Web Design (RWD) to adapt a questionnaire to the continuum of different screen sizes rather than binary optimization.

Notably, the same lead author, Antoun, was more cautious in making a recommendation on approach in another article that reviewed research literature from 2007 to 2016 on smartphone optimization:

> Another issue for SO [smartphone optimized] surveys relates to the design and release of different layouts in response to the size of the respondent's screen or web browser. The discussion so far has focused on adapting a questionnaire designed for PCs into a single optimized version for smartphones; but it is important to note that several studies used a responsive design where several versions of the questionnaires were displayed, with the implicit goal of improving response quality within each version (see, e.g., Amin, 2016). This practice has been adopted from web design where a large number of different visual designs and layouts for a single website are deployed (e.g., for small smartphones, large smartphones, small tablets, large tablets, small desktops, and so forth). What is unclear is whether this approach is also effective for surveys where standardization across layouts is a higher priority. A concern is that variations in a particular layout can affect responses (see, e.g., Smyth, Dillman, Christian, & Stern, 2006). Thus, responsive design calls attention to the need to promote comparability across versions, on the one hand, and to minimize error within each version, on the other.
>
> Whether optimization is binary or responsive to a continuum of screen sizes, another issue is whether to design for the biggest or smallest devices first. Almost all of the reviewed studies started from the point of an existing survey that is designed for PCs and then adapted for smartphones. While this approach may prevent major usability problems in the smartphone survey, it does not necessarily produce an optimal design for smartphones (as the word "optimized" implies). Because the smallest screens appear to pose a greater design challenge, the "mobile first" approach may be desirable to the extent that it does not

[18] Antoun, C., Couper, M., Conrad, F. (2017). Effects of mobile versus PC web on survey response quality. *Public Opinion Quarterly*, Vol 81, Special Issue, 2017, pp. 280-306

have negative effects on the version of the questionnaire displayed on larger browsers (see, e.g., de Bruijne & Wijnant, 2013a; Tharp, 2015). *Future research on the effect of responsive design, with and without a mobile-first design philosophy, is necessary before any firm conclusions on these different approaches are drawn.*[19]

The bottom line is the last sentence: there doesn't seem to be a definitive conclusion yet as to which approach is best.

Questions Addressed by the Panel

The Panel considered the following options with respect to the possibility of a standard encouraging a particular approach to question design/layout across devices.

- *No standard* – e.g. because not enough is known about what approach is best, so leave it up to the researcher to decide what is best for their project
- *Mobile-first* – which uses a mobile-friendly design for all devices, large and small, and minimizes the possibility of "mode" effects
- *Responsive design* – optimize question design/layout for mobile (small) and computer (large) screens, which can lead to different design/layout across devices, but gives the best data quality for each device
- Don't state either of the above as standards, but rather as guidelines on options to consider

Advisory Panel Response

Modern survey platforms have responsive design built in, so responsive design is essentially already the default option.

The Panel's view was that a standard is not appropriate. Reasons include:

- As noted in the background information, the research is not yet settled on which approach would most often be optimal.
- The appropriate option could be different for different surveys, and even for different questions within a survey. That is, for certain questions a responsive design could result in a mode effect – meaning a mobile-first approach would probably be best for those questions, whereas for other questions there may not be mode effects – in which case responsive design for those questions would be optimal.

The Panel said a guideline could be helpful to highlight the options available to a researcher. A guideline to add to Section 2 *Questionnaire Design* could be:

Section 2 guideline:

Design of an online survey should take into consideration the approach taken to the design and presentation of questions on different devices. Options include:

[19] Antoun, C., Katz, J., Argueta, J. and Wang, L. (2017). Design heuristics for effective smartphone questionnaires. *Social Science Computer Review*, Online First

- *Responsive design* - optimize question design/layout for mobile (small) and computer (large) screens, which can lead to different designs/layouts across devices, but gives the best data quality for each device
- *Mobile-first* - which uses a mobile-friendly design for all devices, large and small, and minimizes the possibility of "mode" effects
- *Combination* – Within a survey questionnaire, use responsive design for questions unlikely to be at risk of a mode effect, and use mobile-first for questions where a mode effect is more likely

Background and Questions: Should there be a different survey duration standard for mobile-friendly surveys?

Because completion times tend to be longer on a mobile device than on a computer, a frequent recommendation is that a mobile-friendly questionnaire should be "short." However, there is no consensus on what this means in terms of number of minutes.

The current standard for online questionnaire duration is 20 minutes, but an average duration of 15 minutes or less is "strongly encouraged":

2.1. Standards

1) Survey questionnaires must be designed:
   c) to be completed in a maximum duration of 20 minutes, not including pauses or interruptions. Exceptions could include projects with specialized audiences and those with pre-arranged interviews when the respondent is aware the survey will take longer than 20 minutes. **Average questionnaire durations of 15 minutes or less are strongly encouraged in order to minimize respondent burden.**

Questions Addressed by the Panel

The Panel considered whether the standard for survey duration should be left as is, or revised to specify a shorter duration for mobile-friendly surveys.

Advisory Panel Response

The Panel said the existing Standard on survey duration does not need to be changed. The industry is well aware that survey duration needs to be kept as short as possible, and the Panel agreed that strongy encouraging questionnaire durations of 15 minutes or less is sufficient.

It was also suggested that advances in mobile technology are reducing albeit not eliminating the differences in time to complete a survey. Mobile communication speeds have improved and screen sizes and resolutions have improved to the point where the relative difficulty of completing surveys on a mobile device is compensated for by the convenience of doing a survey when and where a respondent would like.

Background and Question: Should there be guidelines on features of a mobile-friendly questionnaire, and if so what guidelines?

It is difficult to state "standards" – i.e. requirements – for what constitutes a mobile-friendly question design:

- Asking questions that effectively address survey objectives is the top priority, and this may constrain design choices.
- The designs supported by survey software varies across vendors.
- Similarly, software to dynamically adapt question design to device type varies across vendors.
- There is researcher experience and judgment that goes into design choices.

One option in the Standards is not to give any standards or guidelines on mobile-friendly design. The idea is that in *Proposal Documentation* there will be a statement of intent to have a mobile-friendly survey. Beyond that, it will be up to the researcher to decide how to implement that for their survey. And, under this view, guidelines do not belong in a Standards document.

The other option is to give guidelines – i.e. recommended, but not required, practices or principles. This is certainly possible to do based on the research literature. The idea is that it would be helpful to GC POR researchers to have this guidance stated in the Standards document.

For discussion purposes, two approaches for guidance were put to the Panel.

1. The Sage Research report, *Best Practices for Improving Cooperation for Online Surveys*, summarized the following guidelines based on a review of the literature (detailed citations are in the report):

✓ *Write short question text and limit the number of response options.*	The goal is to avoid/minimize scrolling to see the entire question and all of the response options.
✓ *Use large fonts and allocate sufficient space for touch selection of response options.*	This will make it easier to read and respond when using a touch interface, and reduce response errors.
✓ *Avoid requiring zooming to be able to see the question and response options.*	Requiring zooming increases respondent burden, and the time needed to complete the survey.
✓ *Avoid visual clutter – i.e. visual elements that do not directly support efficient presentation of the question and response options.*	These are distracting, and can increase page load times.
✓ *Minimize images, and other high bandwidth requirements.*	Note also that Flash content – which can be used to present video – is not supported on devices using the iOS operating system (iPhones, iPads).
✓ *Use page breaks carefully.*	On the one hand, spreading questions across different pages reduces page load time; on the other hand, too many pages can make for a slow

	experience waiting for pages to load. Orr (2012) suggests a maximum of three questions per page as a rule of thumb; AAPOR (2014) suggests a maximum of two questions per page. However, Mavletova and Couper (2014) found directionally higher break-offs when there was only one question per page compared to multiple questions per page, suggesting it is better to have more than one question per page.
✓ *Consider limiting the requirement for typing/text entry.*	AAPOR (2014) summarizes evidence indicating this has been an issue in the past, but hypothesize that this potential limitation may be going away, as people become more used to texting on their phone.
✓ *Use small grids that require little or no horizontal scrolling, or avoid grids altogether.*	
✓ *Avoid horizontal scrolling; vertical scrolling is more acceptable.*	For example, both Peytchev and Hill (2010), and Stapleton (2013) found that those using a mobile device were more likely to pick the immediately visible response options rather than scroll horizontally to see more options.
✓ *Avoid interactive elements such as sliding scales or drag-and-drop as these are usually too difficult to use on a mobile device.*	

To this list could be added:

✓ *Use survey software that detects the device being used and can automatically adapt question design/layout to the device.*	

2. Another option for guidelines is the five "design heuristics" proposed by Antoun et al (2017)[20] based on their review of the literature:

✓ *Readability*	Text should be large enough to promote easy reading.
✓ *Ease of selection*	Touch targets should be large enough to tap accurately.
✓ *Visibility across the page*	All content should fit the width of the screen so that it is visible without horizontal scrolling.

[20] Antoun, C., Katz, J., Argueta, J. and Wang, L. (2017). Design heuristics for effective smartphone questionnaires. *Social Science Computer Review*, Online First

✓ *Simplicity of design features*	Design features should be simple both for researchers to deploy and for respondents to use. Deploying complicated features (sliders, pickers, questions with embedded video) can increase the occurrence of technical problems. Even without glitches, complicated features can increase missing data or harm data quality if some respondents do not understand or are not able to learn in real time how to use the features. [Note: a "picker" is a rendering for mobile of a drop-down list]
✓ *Predictability across devices* [Comment: We think "Displays appropriately across devices" would be a better label.]	Questionnaires should function in a predictable way across different devices. The concern is that questions will be rendered differently across different types of platforms (e.g., pickers vs. spinners). Another concern involves different screen sizes. Touch targets and font sizes may meet design standards on big devices (e.g., 6-in. displays) but not on smaller devices (e.g., 4-in. displays).

Questions Addressed by the Panel

The Panel considered whether or not there should be guidelines in the Standards on what makes a questionnaire mobile-friendly.

Advisory Panel Response

The Panel said there should not be any required features (i.e. standards) for mobile-friendly questionnaires. Reasons included:

- There is a subjective element to many of the features – e.g. what is a "large font"?
- Advances in mobile technology – including processing power, network speed and screen resolution – are changing what can be mobile-friendly.

One view was that the Standards should not address this topic at all, but the majority felt that a list of examples, in the form of a guideline, could be a useful reminder to researchers.

The following would be a possibility to include in Section 2 *Questionnaire Design*:

Section 2

The following are examples of practices that can make a questionnaire more mobile-friendly:

1. Avoid horizontal scrolling
2. Avoid interactive elements that would be difficult to use on a mobile device
3. Avoid large grids, or avoid grids altogether
4. Minimize visual clutter
5. Allocate sufficient space for touch selection of response options
6. Write short questions and limit the number of response options
7. Minimize images and other high bandwidth requirements

Proposed Revisions to *Pre-testing* in the Online Standards

Background and Questions

There are two types of pre-testing that can be done for an online survey questionnaire where some respondents may use a mobile device while others may use a computer:

1) Internal pre-testing by the researchers to see whether the questions display appropriately on some of the different devices that respondents might use
2) External pre-testing with respondents using different devices

The current standards do not specifically refer to requirements for online surveys where the survey is being completed on both computers and mobile devices.

In Section 3 of the Online Standards, the two most relevant current standards are:

3. PRE-TESTING

3.1. Standards

1) Pre-testing of all components of a new or revised survey questionnaire that may influence data quality and respondent behaviour is required. This includes the online appearance and functionality of the questionnaire.

5) A minimum of 10 pre-test surveys are to be completed in each language in which the final survey will be fielded. An exception could be projects with small survey populations, such as a client-satisfaction survey of a small client base. In such cases the researcher must, in consultation with the client, take steps to ensure that the smaller number of pre-tests are sufficient to guarantee questionnaire quality. For example, a cognitive pre-test may be warranted.

Questions Addressed by the Panel

The Panel considered the following questions:

Should standards specific to completion of an online survey on mobile devices be added?

An argument for not adding anything is that Standard 3.1.1 is sufficiently broad in scope that a requirement for pre-testing of the mobile-friendly aspects of the survey is implied even if not explicitly stated.

An argument in favour of adding standards is to make sure the pre-testing is done appropriately – with the added standards stating the minimum requirements.

What, if any, revision should be made to Standard 3.1.1 in *Pre-testing*?

For example, a revision could be to add a requirement for some internal pre-testing on different devices:

1) Pre-testing of all components of a new or revised survey questionnaire that may influence data quality and respondent behaviour is required. This includes the online appearance and functionality of the questionnaire.

    a) For mobile-friendly surveys, researchers must also do internal pre-testing on a sample of the types and sizes of devices that respondents might use. On smartphones, the internal pre-testing must look at how questions display in both portrait and landscape mode.

What, if any, revision should be made to Standard 3.1.5 in *Pre-testing*?

For example, a revision could be to require pre-testing on a sample of different devices:

5) A minimum of 10 pre-test surveys are to be completed in each language in which the final survey will be fielded. An exception could be projects with small survey populations, such as a client-satisfaction survey of a small client base. In such cases the researcher must, in consultation with the client, take steps to ensure that the smaller number of pre-tests are sufficient to guarantee questionnaire quality. For example, a cognitive pre-test may be warranted.

    a) For online surveys the pre-test must include respondents using at least some of the different device types and sizes that might be used to complete the survey. For example, in a survey that can be completed on either a mobile device or a computer, the sample must include a specified minimum number of both mobile device users and computer users.

Advisory Panel Response

Section 3.1.1 of the *Pre-testing* standards should be revised for online surveys completed on both computers and mobile devices to include a requirement to pre-test on different devices.

Section 3.1.5, which requires a minimum of 10 pre-test interviews in each language, should be left as is.

The Panel was split in terms of how 3.1.1 should be revised to include a reference to pre-testing on different device types, with more favouring Option 1 below than Option 2. Option 1 adds a requirement for internal pre-testing on different device types, but leaves the existing external pre-testing requirement as is – i.e. with no requirement to externally pre-test on different device types. Option 2 adds both internal and external pre-testing requirements for different device types.

Option 1: Require internal pre-testing on different devices but not pre-testing with respondents using different devices (external pre-testing): Some degree of internal pre-testing by the research company is a reasonable requirement, but routinely requiring external pre-testing can be impractical and is not usually necessary if the internal pre-testing results are satisfactory.

An issue with mandating an external pre-testing requirement is that the researcher does not know in advance the type of device a respondent will use. Trying to complete a specified number of pre-tests for different device types can substantially increase the total number of people who have to be contacted to complete the minimum quotas by device type.

There were different views on how to express the scope of an internal pre-testing requirement, including:

- Internally pre-test on at least one smartphone and at least one computer/tablet.

- Internally pre-test on a sample of the types of devices that respondents might use.

  There was some discussion of whether or not size of smartphone screen should also be referenced, however this is generally not necessary because if a question displays properly on one size of smartphone screen it very likely displays properly on other smartphone screen sizes.
- Internally pre-test across all technologies, including smartphones, laptops, tablets, desktops/PCs. For smartphones, pre-testing should be required for both Apple and Android operating systems and on current, common smartphone types.

There was also a difference of opinion as to whether or not the internal pre-testing should require pre-testing in both portrait and landscape modes, although most did not feel that pre-testing in both modes should be required. People who use smartphones are familiar with changing the phone orientation should they feel a need.

Option 2: Require both internal and external pre-testing on different devices: The wording for the external pre-test could be:

> For online surveys, the pre-test must include respondents using at least some of the different device types that might be used to complete the survey. For example, in a survey that can be completed on either a mobile device or a computer, the sample must include a specified minimum number of both mobile device users and computer users.

Note that it was not suggested to increase the current required minimum number of pre-tests (10 in each language), so the quotas by device types would be within the current required minimum.

Possible Revisions to *Data Collection* and *Quality Controls* in the Online Standards Related to the Possibility of Mode Effects by Device Type/Screen Size

Background and Questions

In a survey that allows completion on both mobile devices and computers, and particularly one using a "responsive design" approach that can result in different question designs/layouts for different size screens, there is the potential for a "mode" effect. That is, the different designs/layouts for a given question could cause different response distributions. If this possibility is to be explored, then data on device type needs to be collected during the survey.

Also, there is the possibility of a device-type effect as a result of differences in the characteristics of people who use a mobile device to complete a survey versus people who use a computer. For example, a research vendor noted the following in an article advocating for "device agnostic" sampling for online surveys:

> Specific pockets of the population gravitate toward mobile and we expect to see the level of systematic non-coverage bias to grow in non-mobile research designs. Some of our testing demonstrates between 20-25 percent of millennials (those born from 1981-2000) prefer to access surveys via mobile, which means our non-mobile surveys are missing the views of a substantial portion of this audience. Considering that this group is likely more tech savvy

and connected with peers, we expect over time to see biased and inaccurate results when they are excluded from our sampling frame.[21]

This illustrates the importance of a mobile-friendly survey for ensuring good coverage of the population. However, it also means that an exploration of possible effects of question design/layout would need to be done carefully to avoid confounding question design/layout with covariates such as age.

The conclusions from the research literature are:

- There tend not to be significant "mode" effects.
- But, the research literature to date is inconclusive because of the limited number of studies and the limited question design variations studied.

Questions Addressed by the Panel

The Panel considered whether or not there should be any requirement to collect information on device type, and any requirement to conduct an analysis for mode effects by device type. With regard to the latter, the following was suggested to the Panel for purposes of discussion:

14.6. Quality Controls

4) For online surveys where there are substantial differences in question design/layout by device type, detailed description of any data quality issues arising from combining data collected from the different question designs/layouts.

Advisory Panel Response

A majority of Panelists supported having a standard requiring collection of device information, but did not support having any standard requiring an analysis of mode effects for each survey. This view was paired with a recommendation that "research on research" be done using the aggregated device data collected across surveys to determine what if any standard would be appropriate for analysis of potential mode effects. The view was that not enough is known about device type mode effects at this time to specify analytic requirements for individual surveys, and it might turn out that really there is not a problem that needs to be addressed in the Standards. It was noted that if device information is collected, the researcher has the option to explore mode effects if that is judged to be useful for a particular survey.

Under this view, the following standard could be added to Section 7.7 *Monitoring of Online Survey Fieldwork:*

7.7. Monitoring of Online Survey Fieldwork

3 If possible, the survey system should collect data to allow classification of a device as a smartphone, tablet or computer. If this is not possible, there must be a question in the survey to gather this information.

Several Panelists had a different take on possible requirements for collection of device type information and analysis of potential mode effects by device type:

[21] Simpson, S. (2014). Okay fine, I'll go device agnostic. *Survey Magazine*, June 2014.

- One Panelist favoured having a standard requiring both collection of *device information* and an analytics standard for each survey such as the one suggested for discussion purposes – i.e. "detailed description of any data quality issues arising from combining data collected from the different question designs/layouts."
- A few Panelists said neither type of requirement should be incorporated into the Standards, at least not until there is more conclusive research on what if any standards might be appropriate. In addition to the uncertainty over what would constitute appropriate standards, considerations include:
  - Not all survey platforms collect device data, or the same types of device data. In this context, adding a question to get this information – while perhaps having some value – has to be weighed against keeping questionnaire duration as short as possible and ensuring that all substantive questions are asked.
  - If a survey has been designed to have a mobile-friendly version of the questionnaire, and there has been internal pre-testing on different devices, the data collection and analytic standards are not necessary.
  - Device type usage can be strongly correlated with other variables – in particular, age – which can make it very difficult to do conclusive analysis of mode effects, at least at the level of individual surveys.

Use of Mobile Devices – Covering Respondent Costs

Background and Questions

Users of mobile devices may incur costs to participate in a research survey.

The current standards do not have any requirements as to how such costs should be handled.

The following is a standard/guideline in the *ESOMAR/GRBN Guideline on Mobile Research:*

> **3.1.3 Costs**
>
> Unlike most other research methods, data subjects may incur costs as a consequence of participating in mobile research that may include charges for data downloads, online access, text messaging, data plan overages, roaming charges, voicemail message retrieval and standard telephone charges. Researchers should design their research so that data subjects incur no costs without express approval. If this is not possible, researchers must be prepared to offer compensation. Such compensation may be cash, mobile money, airtime or other forms of value.

Note the ESOMAR/GRBN explain their use of "must" and "should" as follows:

> Throughout this document the word "must" is used to identify mandatory requirements. We use the word "must" when describing a principle or practice that researchers are obliged to follow. The word "should" is used when describing implementation. This usage is meant to recognise that researchers may choose to implement a principle or practice in different ways depending on the design of their research.

Questions Addressed by the Panel

The Panel considered whether something like the ESOMAR/GRBN 3.1.3 above should be incorporated into the Online and Telephone Standards.

Advisory Panel Response

The Panel's view was that there should not be a standard about covering respondent costs associated with using a mobile device, and several reasons were given.

Respondents always have a choice whether or not to participate in a GC POR online or telephone survey. There was some discussion of what to do in the case of a survey where respondents are required to use a mobile device. However, even in this case respondents can choose not to do the survey.

The current standards require certain information be given about the survey (e.g. length), so respondents are able to make an informed choice about whether or not to participate.

It is reasonable to assume that mobile device users know that there are costs to using a mobile device, albeit what the cost is for participating in a particular survey can vary considerably, and may not even be known, or could only be calculated in some arbitrary fashion (e.g. what would the cost be if a person has an unlimited data plan). Since everyone knows at least roughly what their mobile plan and costs are, it is reasonable to assume this is factored into their decision whether or not to participate in a survey.

Unless compensation is set at an arbitrary fixed amount for all mobile users, the logistics of determining the amount to compensate each respondent and documenting this for billing purposes would be very complex and difficult, if not impossible. There is a wide variety of data and voice plans on the market, at different prices. When a plan provides a certain number of minutes or gigabytes, calculating a cost for a survey completed within those parameters would be largely arbitrary, and very difficult to document. Overall, the cost of administration of compensation would be considerable.

In the absence of a standard, research firms would still have the option to address compensation should that prove useful for a particular survey. For example, if response rate is a concern for a particular survey, respondents could be offered an option to complete the survey at a different time or on a computer, or offered a fixed incentive. However, the need for these sorts of actions is best determined on a case by case basis, and in consultation with the GC client concerning cost recovery of such incentives.

Inclusion of Cell Phones and Landline Phones in Telephone Surveys

An important issue in sampling for telephone surveys is the inclusion of cell phone users and landline users. This can affect coverage of the survey population, the sampling frame(s) used for the survey, and possibly weighting.

The following are some statistics taken from the CRTC's *Communications Monitoring Report 2017*:

- The data are from Statistics Canada's 2015 Survey of Household Spending, so these are outdated. Very likely, the percentage of cell phone only has increased since then, and landline-only and dual ownership has decreased.
- Cell phone only households in Canada = 27.5%, landline-only = 13.2%, and dual cell-phone/landline = 58.6%
- The types of phones in households varies demographically:
  - -- The percentage of cell phone only households varies by household income, being highest in the lower-income households: lowest income quintile = 35.0%; highest income quintile = 18.8%
  - -- The percentage of cell phone only households varies by province, from 14.4% in New Brunswick to 34.7% in Alberta.

  The CRTC report does not report phone type by age, but back in 2010, the AAPOR Cell Phone Task Force concluded: *"young adults in the U.S. aged 18 to 34 years, can no longer be reached successfully via the landline frame."*[22]

These data indicate that a telephone probability sample of the general Canadian adult population must include a cell phone sample.

The PORD literature review for this project noted the following:

- No industry associations have set standards for numeric proportions of landline and cell phone numbers.

  As an example of what a respected survey organizations does currently in the U.S., the literature review cites Pew Research, which now targets to complete 75% of interviews by cell phone and 25% by landline[23]. The literature review notes, though, that *"cellphone-only households have not reached the same level of penetration in Canada as compared to the United States."*

- The optimal proportion of cell phones versus landline phones will vary depending on the population being surveyed: *"The optimal allocation for a dual-frame telephone survey depends*

[22] AAPOR Cell Phone Task Force. (2010). *New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. with Respondents Reached via Cell Phone Numbers.* This is a very good review of basic issues that need to be considered when including a cell phone sample in a telephone survey.

[23] Pew Research, *Our Survey Methodology in Detail*, http://www.pewresearch.org/methodology/u-s-survey-research/our-survey-methodology-in-detail/

on the proportion of the population that only has a cellphone and the target audience. … For example, it will be more cost effective to conduct a survey of young adults almost exclusively using cellphone sample…"

- It is premature to use only a cell phone sample frame to get a good probability sample in Canada. Rather, it is necessary to use a dual frame design – i.e. both a cell phone sample frame and a landline sample frame.
- With regard to weighting a dual-frame sample: *"At the time of writing, there is no consensus on the best approach to weighting dual-frame survey samples."*

  The literature review suggests that in Canada, often the following approach is used:

  > In Canada, when conducting a telephone survey of the general public, it is common to apply one set of demographic weights to survey data thereby treating cellphone and landline interviews the same. A dual-sample frame is used to minimize the potential for coverage error resulting from the incomplete landline sample frame … From this perspective, survey samples resulting from dual-frame sampling only require weighting adjustments to bring the demographic distributions in line with those of the population. (p.14)

- The literature review recommends: (a) don't set numeric standards for the proportion of cell phone and landline numbers, and (b) have a "guideline" (not a standard/requirement) that survey data reflect the population of cell phone only households:

  > Sample frame proportions for dual-frame surveys: The proportion of landline and cellphone numbers included in dual-frame samples varies as researchers try to find the optimal balance and as the incidence of exclusive cellphone use increases. For this reason, it is recommended that PORD not put in place prescriptive standards for dual-frame sample proportions. There will not be one ratio of landline-to-cell sample that will apply to all public opinion research studies. Other factors, such as budget, target audience, and field period, will influence the proportions.
  >
  > A useful guideline would be one that focusses on the outcome of the data collection. That is, the sample ratios chosen should ensure that the proportion of cellphone-only households in the final survey sample reflects that of the population of cellphone-only households at the time of the study.

The 2010 AAPOR Cell Phone Task Force noted a variety of other issues involving surveying cell phone users, and typically noted that there is no definitive "best" resolution. For example:

- How to deal with the fact that a cell phone sample frame overlaps with a landline sample frame (i.e. dual users are in both)
- How to handle dual users who are "cell phone mostly" or "landline mostly"
- Selection of respondent in a household when dialing a cell phone (e.g. Pew Research selects the person answering the cell phone, but for landline has a process for selecting someone in the household who may not be the person answering the phone)
- Selection of respondent when there is cell phone sharing within a household
- How to treat a business cell phone that also may be used sometimes for personal matters

The following are the current standards in the Telephone Standards that refer to cell phones:

1. PROPOSAL DOCUMENTATION

1.2. Technical Specifications of the Research

1.2.3. Response Rate/Participation Rate and Error Rate

1) State the expected response rate/participation rate for the sample(s). State expected response rates/participation rates for each mode when more than one mode is proposed (i.e. separately for online and telephone components). Similarly, if a component of the research will recruit respondents via their **cell phones**, the expected response rate/participation rate for that component must be stated separately as well. The expected response rate/participation rate is an estimation based on various factors such as previous response rates/participation rates, trends etc. The proposal must include a brief discussion of the factors that might cause the actual response/participation rate to fall short of the stated target.

1.2.4. Description of Data Collection

7) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method. When it is proposed to recruit at least some respondents via their **cell phones**, provide a rationale for using this approach in addition to or in place of landline phones.

4. SAMPLING PROCEDURES

4.2. Probability Sampling

3) A full description of the sample design and selection procedures must be stated including:

   c) at each sampling stage, the method of achieving a probability sample (e.g., random selection) shall be explained, and any subsets of the universe that have been excluded or underrepresented shall be stated (e.g., **cell phone only** households), although whenever possible, an estimate of the percentage of the universe that has been excluded or underrepresented must be provided;

Note that Section 1.2.4 is repeated in Section 15.5 *Data Collection* (telephone) as part of Section 15, *Mandatory Survey Report Requirements*.

Revisions to Section 1.2.3 #1 in the Telephone Standards

Background and Questions

Questions Addressed by the Panel

The Panel was asked to consider whether there should be any revisions to Standard 1.2.3 #1 in *Proposal Documentation* in the Telephone Standards, which states requirements to provide information about estimated response/participation rates. Currently it requires an overall estimated response/participation rate, and if relevant an estimated response rate for cell phones:

1. **PROPOSAL DOCUMENTATION**
   1.2. **Technical Specifications of the Research**
      1.2.3. **Response Rate/Participation Rate and Error Rate**
      1) State the expected response rate/participation rate for the sample(s). State expected response rates/participation rates for each mode when more than one mode is proposed (i.e. separately for online and telephone components). Similarly, if a component of the research will recruit respondents via their cell phones, the expected response rate/participation rate for that component must be stated separately as well. The expected response rate/participation rate is an estimation based on various factors such as previous response rates/participation rates, trends etc. The proposal must include a brief discussion of the factors that might cause the actual response/participation rate to fall short of the stated target.

Advisory Panel Response

Some Panelists said Section 1.2.3 is fine as is, while others said for telephone surveys it should require the estimated response rate, if relevant, for both cell phones and landlines. An argument in favour of the latter is that telephone surveys typically include both a cell phone sample and a landline sample.

Response/participation rate estimates are useful at the proposal stage because they are an important driver of the planned level of fieldwork effort needed to achieve the desired sample size.

The Standards appropriately do not require reporting the actual response/participation rates achieved for cell phones versus landlines. Calculation of the response/participation rates actually achieved for cell phones versus landlines is subject to error. This is because there is some uncertainty over how to classify phone numbers when an interview is not completed. The sample used for dialing will usually be divided into a cell phone sample and a landline sample, but these classifications are imperfect due to number porting and the existence of exchanges which contain a mix of cell phone and landline numbers.

Revisions to Section 1.2.4 #7 in the Telephone Standards

Background and Questions

Questions Addressed by the Panel

This standard currently states:

1. **PROPOSAL DOCUMENTATION**
   1.2. **Technical Specifications of the Research**
      1.2.4. **Description of Data Collection**
      7) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method. When it is proposed to recruit at least some respondents via their cell phones, provide a rationale for using this approach in addition to or in place of landline phones.

This standard states that a rationale must be given when the sample includes interviews on cell phones. The language arguably overly downplays the importance of including cell phone users in the sample. For purposes of discussion, the following revision was suggested to the Panel (leaving out the multi-mode component, which was not relevant to the discussion):

> **1.2.4 #7:** The telephone survey should (must?) include interviewing by both cell phone and landline. The sample ratios of cell phone to landline must be stated, and should (must?) ensure that the proportion of cell phone only households in the final survey sample reflects that of the population of cell phone only households at the time of the study.
>
> If interviewing is to be done only by landline or only by cell phone, provide the rationale.

Advisory Panel Response

The suggested text for Section 1.2.4 #7 does not contain a specific recommended numeric ratio for cell phone sample to landline sample, nor should it. There is not good, up-to-date information that would support stating a specific numeric ratio.

For the same reason, the Panel deleted the phrase, *ensure that the proportion of cell phone only household in the final survey sample reflects that of the population of cell phone only household at the time of the study*. Again, there is not always good, up-to-date information to make this possible. Further, even if the ratio is known, it would be onerous and expensive to ensure the ratio for population segments or geographic zones more narrow or granular than at the national and provincial/territorial levels.

The Panel was split on how to revise Section 1.2.4 #7. There were three different positions expressed with respect to how to revise the section:

- *Sampling both cell phones and landlines should be a guideline, not a standard*

  This recognizes that sampling both cell phones and landlines is the current common practice, but also allows for budget flexibility. The cost per interview is higher for cell phone sample, and including a cell phone sample may exceed a client's budget, or may be limited by a research firm's current menu pricing on the GC POR Standing Offer.

- *Sampling both cell phones and landlines is required, and a sampling ratio and rationale must be stated*

  The following is an example of how this view could be worded:

  > **1.2.4 #7:** Telephone surveys with the general population or most its subgroups must include interviewing by both cell phone and landline. The ratio of cell phone sample to landline sample must be stated, and the rationale for the ratio must also be provided. If interviewing is to be done only by landline or only by cell phone, provide the rationale.

  Because of the large and growing importance of cell phone usage, the view is that including both cell phone and landlines in the sample should be the default requirement.

  The phrase *"with the general population or most of its subgroups"* is there both to make the point that interviewing by both cell phone and landline is widely applicable, but also to

acknowledge there may be subgroups where cell phone only or landline only would be appropriate.

The research firm should give a reason for its recommended sampling ratio, but recognizing that this will likely be based on imperfect population data.

- *Sampling both cell phones and landlines is required, and a sampling ratio must be stated – but not a rationale for the ratio*

  The following is an example of how this view could be worded:

  > **1.2.4 #7:** Telephone surveys with the general population or most of its subgroups must include interviewing by both cell phone and landline. The ratio of cell phone sample to landline sample must be stated. If interviewing is to be done only by landline or only be cell phone, provide the rationale.

  The reasons for dropping the requirement for a rationale are (a) lack of good, up to date population data, and (b) even with population estimates, a different ratio could be used and be weighted back to population ratio.

Note that the versions of Section 1.2.4 #7 above use the phrase "sampling ratio" and refer to the the "cell phone sample" and the "landline sample." As a proposal documentation requirement, this language focusses on the ratio of cell phone sample that will be dialed versus landline sample that will be dialed. The ratio does not directly apply to the actual ratio of completed interviews for cell phones versus landlines, which could end up being somewhat different. A Panelist suggested an alternative approach is to frame the ratio stated in Section 1.2.4 #7 as being what will be achieved in the completed interviews. From this perspective, the language used in Section 1.2.4. #7 should be something like the following: *The ratio of interviews to be completed by cell phone versus by landline must be stated.* The argument for this alternative is that the ratio of the completed interviews states what the client can expect in the completed unweighted data set, rather than the ratio used at the intermediate step of the sample of numbers to be dialed. The choice between these two alternatives is significant, as it affects the composition of the sample of numbers to be dialed and quota controls on completion by device type during fieldwork.

Sampling Procedures, Standard 4.2.3c

Background and Questions

Questions Addressed by the Panel

The current Standard 4.2.3c in *Sampling Procedures* addresses disclosure of coverage issues in probability samples, and gives as an example a sample of cell phone only households. The question was whether or not also to include an example of landline-only households.

Advisory Panel Response

The Panel recommends adding landline-only samples as another example. Given the growing number of cell phone only households, a landline-only sample could have substantial coverage error.

4. SAMPLING PROCEDURES

4.2. Probability Sampling

3) A full description of the sample design and selection procedures must be stated including:

c) at each sampling stage, the method of achieving a probability sample (e.g., random selection) shall be explained, and any subsets of the universe that have been excluded or underrepresented shall be stated (e.g., cell phone only or landline only households), although whenever possible, an estimate of the percentage of the universe that has been excluded or underrepresented must be provided;

Telephone Survey Call-back Requirements

Background and Questions

The Telephone Standards for call-backs are in Section 7 *Data Collection*:

> **7.2. Call-Backs**
>
> 1) There will be a minimum of eight call-backs made before retiring a telephone number and substituting it with another number. The call backs must be made at varying days and times over a minimum seven-day period. An exception could be made when the field period is shorter as a result of the need to assess recall of particular events or activities.
> 2) Every effort must be made to ensure that the respondent is called back if an appointment has been arranged and that the date and time of that appointment are respected.
> 3) No attempt will be made to call back refusals.

PORD posed a question to the MRIA about the appropriate number of call-backs, as concern had been expressed about whether eight call-backs is too many, and might be perceived as harassment.

The MRIA *Polling Standards for the Canadian Marketplace*[24] state eight call attempts as a maximum, as compared to the Telephone Standards Section 7.2 *Call-backs* which states eight call-backs as a minimum. Note that the MRIA and Section 7.2 differ in two ways: (1) MRIA specifies a maximum whereas Section 7.2 specifies a minimum, and (2) MRIA specifies call attempts whereas Section 7.2 specifies call-backs. The term "call-back" implies there is an initial contact attempt followed by call-backs. Rephrasing the rule in Section 7.2 in terms of call attempts, it is stating that there should be a minimum of 9 call attempts.

The MRIA *Polling Standards* also state a definition of call-backs (no definition is given in Section 7.2 of the Telephone Standards).

- Make no more than eight (8) calls to the same telephone number. This number includes:
  - callbacks made to establish initial contact with the potential respondent (for instance, when there was no answer or the line was busy on previous calls); and
  - callbacks made after contact has been established (for instance, when the potential respondent asked for a callback at a more convenient time, or when the selected respondent was not home)

MRIA posed the following question to some members: *The MRIA had adopted the policy of a maximum of 8 calls to each potential respondent. See Appendix "L" 8.4.2 . If the number of calls allowed are reduced would this affect your research studies? What is your firm's frequency of callbacks - Over what period do you make the (eight) 8 calls?* The response relayed to PORD was:

> Members are very sensitive to respondent fatigue and aim to regulate the frequency of calls to the same respondents.

[24] MRIA Code of Conduct for Market and Social Research, Appendix L, *Polling Standards for the Canadian Marketplace*

The additional calls to respondents are usually made] after exhausting the list of potential respondents and calls are made to those whom the researcher was unable to contact in the initial call.

It is extremely important to recognize the difference between dispositions for call attempts. Not every call attempt should be considered a call-back. For example, 8 call backs, all yielding a busy signal, is very different than 8 call backs all resulting in an answer and call back request. At the same time, the number of call backs must be large enough to provide a reasonable expectation of equal probability of selection for all primary sample units. We have a maximum of 7 call backs, but do consider some call dispositions to be partial callbacks (e.g., a busy signal counts as 1/3 of a call-back). So theoretically we could call a phone number up to 21 times (21 busy signals). Usually phone numbers are resolved after 6 - 10 attempts.

Refusal conversion dialing must also be considered and whether these conversion attempts are considered within the call back limit.

The Panel was also asked to comment on whether the call-back requirements should be the same for respondents using a cell phone. The MRIA *Polling Standards* do not specify different call-back requirements for respondents using a cell phone versus a landline. However, the MRIA *Framework for Live Telephone Standards*[25] states that consideration should be given to making a smaller number of call attempts to respondents using a cell phone:

MRIA Appendix P – *Framework for Live Telephone Standards*

Section 6 – Special Treatment When Dialing Cell Phone Banks – Minimum Standard

6. Organizations should consider whether a more modest maximum for call attempts should be used for cell phones rather than landlines

Questions Addressed by the Panel

The questions considered were:

- What should be said in the Standards about the number call-backs, including whether there should be a different standard for respondents reached on a cell phone
- Whether or not the Standards should state a definition of call-backs

Advisory Panel Response

Number of Call-backs

With regard to terminology, there is a difference between "call-backs" and "call attempts": the implication is that "call attempts" equals one plus the number of "call-backs." There is also some ambiguity in the meaning of "call-back", in that it can be interpreted in a general way, or more narrowly when the person answering the phone has requested to be called back. The Panel's recommendation is that the updated Standards refer to "call attempts" rather than "call backs" on the grounds that the meaning is more straightforward and it aligns with the terminology used in the MRIA *Code of Conduct*.

[25] Marketing Research and Intelligence Association (MRIA), *MRIA Code of Conduct for Market and Social Media Research – Appendix P: Framework for Live Telephone Standards – Guideline for Conducting Live Telephone Market Research*

The Panel's view is that a minimum of eight call-backs (nine call attempts) is excessive, and a smaller number should be specified.

The following were considerations in the discussion:

- *Cell phone versus home phone/landline:* On the one hand, there is an argument in favour of doing fewer call attempts to cell phones, because these are personal devices that people usually carry around with them. Therefore, a person's availability to take a call when the phone rings tends to be higher with a cell phone. On the other hand, the type of phone a respondent will be using is not known with 100% certainty, and it is logistically simpler to apply the same call-back standard to all phone numbers. The Panel opted to apply the same call-back requirement to both cell phones and home phones.
- *Fieldwork duration:* Fieldwork duration can place a practical limit on the number of call-backs that can be done. This is already acknowledged in the current standard, as it says, *an exception could be made when the field period is shorter…"* The Panel opted to retain this language, and not to specify different call-back requirements as a function of fieldwork duration.
- *Minimum, maximum or both:* From a standards perspective, these serve different purposes:
  - *Minimum:* The focus here is on data quality, to ensure a reasonable effort is made to contact a sample unit, and thereby avoid potential bias (in case people contacted after one try are different in some relevant way from people contacted after multiple tries).
  - *Maximum:* The focus here is on avoidance of harassment. There is also an issue of diminishing returns, which is more about fieldwork efficiency. Experience has been that the return in terms of response rate improvement diminishes quickly after the first several call-backs.

  One Panelist suggested the Standards should specify both a minimum and a maximum (numeric values were not specified), while most suggested specifying only a minimum number of call-backs. A reason for the latter is that research firms do not have a default urge to do a large number of call-backs. Call-backs cost money and time, there are rapidly diminishing returns for response rate, and firms do not want to harass people. Research firms are motivated, for good reasons, to limit the number of call-backs. Further, there can be circumstances when a larger number of call-backs is warranted, such as a limited number of sample units, or a hard-to-reach target population. Specifying only a minimum number of call-backs gives research firms more flexibility to adjust the number of call-backs upwards if or as needed.

The majority of Panelists concluded that the appropriate minimum number of call attempts is six (meaning five call-backs), while several opted for a minimum of five call attempts (meaning four call-backs).

There was some discussion about call-backs after an appointment has been made. This is in reference to Section 7.2.3, *Every effort must be made to ensure that the respondent is called back if an appointment has been arranged and that the date and time of the appointment are respected.* A few Panelists suggested specifying that at least two call-backs be made to reach the respondent. Some other Panelists said this level of specificity is not needed for this situation because the

research firm would already likely be motivated to make “every effort” to interview someone who has expressed interest in participating.

The Panel’s view was that it is not necessary to put a definition of call attempts in the Standards. In this context, there was discussion of whether or not a busy signal should be treated as a call attempt, and there was disagreement on this point. There were two alternative points of view:

- *Busy signal is a call attempt:* In literal terms, dialing and getting a busy signal is a call attempt. Another argument was that it is possible, now or in the future, for phone companies to provide a service in which a busy signal is played for call screening or call blocking purposes.
- *Busy signal is not a call attempt:* A respondent may not be aware they were called, so re-dialing would not constitute harassment. A research firm’s dialing protocol may be to re-dial two or three times within a short period of time after getting a busy signal.

Because the standard specifies only a minimum number of call-backs, variability across research firms in treatment of busy signals is not an issue: some will treat it as a call attempt and perhaps end up closer to the minimum number of call attempts, while others will not treat it as a call attempt and perhaps end up doing more call attempts. Since busy signals are unobtrusive for respondents, the extra dialing would not risk harassment.

Interactive Voice Response (IVR) Telephone Surveys

In an Interactive Voice Response (IVR) telephone survey, a computer is programmed with a questionnaire, calls are made automatically, and the recorded questions are read by the computer. There are no live interviewers.

The literature review commissioned by PORD described the following advantages and disadvantages of IVR surveys:

Advantages of IVR

- More time and cost effective compared to live interviewer surveys
- Because of its lower cost, IVR can be useful for surveying low incidence populations
- Less subject to social desirability effects
- Have done well in some election forecasting surveys in Canada

Disadvantages of IVR

- Works less well for lengthy surveys due to hang-ups, so it is better suited to short surveys
- Questions about sample quality: the literature review specifically mentions an IVR design in which there are no call-backs, and the *"survey sample will include whoever could be reached on the night of the survey, which introduces the possibility of non-response bias."* The review notes that there is no guarantee that weighting will eliminate non-response bias.
- Negative perceptions of the methodology, which may be particularly important for Government of Canada public opinion research

Section 5.3.1 Use of IVR

Background and Questions

Section 5.3 *Use of Interactive Voice Response* in the Telephone Standards discourages, but does not forbid, use of IVR surveys for POR. It also suggests circumstances when IVR may be an appropriate methodology. The standard states that IVR surveys have the same requirements as interviewer surveys for the survey introduction, respondent opt-out, times when calls can be made, and delay in acknowledging an answered call.

5. RETAINING PUBLIC CONFIDENCE

5.3. Use of Interactive Voice Response

1) Characteristics of Interactive Voice Response (IVR) surveys, including the impersonal style conveyed by automation, put that method in conflict with the manner in which the Government of Canada wishes to engage Canadians. IVR can therefore be used only when a convincing case is made that the specific information to be collected is essential for making important decisions and cannot be obtained through other means. For example, IVR may be judged acceptable when the opinions of a hard-to-reach (low incidence) group are critical to the issue at hand and the very high call volume made economical by IVR is likely to markedly

increase participation from members of that group in the survey. (IVR may also be used whenever respondents have agreed beforehand to this method.)

2) When IVR is used, the same information required for interviews conducted by live interviewers (sponsor, researcher, participation is voluntary, assurance of confidentiality, etc.) must be included in the IVR survey introduction. Respondents must also be provided early in the introduction with an easy method to opt out of the survey (e.g., by pressing a specific key) so that the call is terminated gracefully and no more calls are made to that number. The same requirements for the time-of-day of calls and delay in acknowledging an answered call (see section 5.2.) apply to the use of IVR.

Questions Addressed by the Panel

The Panel considered whether there should be changes to Section 5.3.1 *Use of Interactive Voice Response*.

Advisory Panel Response

Within the Panel, there was a range of views on the use of IVR for GC POR surveys, including:

- IVR surveys are rarely appropriate for GC POR surveys: Response rates are low, and consistent with this it appears people do not like getting or responding to surveys via IVR. Section 5.3 should take a stronger stand in discouraging use of IVR.
- The language in Section 5.3 is overly and unnecessarily prejudicial against IVR surveys: Like any other survey methodology, IVR surveys have pros and cons, and its possible use for a project should be considered on this basis just like any other survey methodology.

The majority of Panelists were somewhere in between these two views.

Specific considerations pertinent to revising Section 5.3.1 included:

- IVR can be an efficient and cost-effective method for time-sensitive surveys where the field duration is very short.
- IVR can be used in conjunction with an interviewer, where there is switching between an interviewer and the IVR system. For example, an interviewer could initiate the survey, then at some point transfer the respondent to an IVR system for some questions, such as questions on sensitive topics where there could be fewer social desirability response biases on IVR compared to talking with an interviewer. IVR can also be used in a multi-phase study, for example when IVR is used as a recruiting tool for respondents to complete a survey with an interviewer or online.
- There may be CRTC regulations which apply to IVR surveys and which could require modifications to Section 5.3.2. These regulations may affect when IVR survey calls can be made, and information disclosure requirements. There was some uncertainty about this, and PORD should determine what, if any, CRTC regulations affect IVR surveys and incorporate this into Section 5.3.

Based on the considerations above, the majority of Panelists supported the following modification to Section 5.3.1 (subject to further modification based on CRTC regulations):

5.3. Use of Interactive Voice Response

1) Characteristics of Interactive Voice Response (IVR) surveys, including the impersonal style conveyed by automation, put that method in conflict with the manner in which the Government of Canada wishes to engage Canadians. IVR can therefore be used only when a convincing case is made that the specific information to be collected is essential for making important decisions and cannot be obtained through other means. For example:
   - IVR may be judged acceptable when the opinions of a hard-to-reach (low incidence) group are critical to the issue at hand and the very high call volume made economical by IVR is likely to markedly increase participation from members of that group in the survey.
   - IVR may be judged acceptable for time sensitive surveys where the field duration is very short.
   - IVR may be judged acceptable when it is part of a mixed-mode design, for example when respondents identified through another survey mode are directed to an IVR, or a design in which there is switching between a live interviewer and IVR.
   - IVR may also be used whenever respondents have agreed beforehand to this method.
2) [See next section]
3) The IVR survey must follow the relevant regulations in the CRTC's Unsolicited Telecommunications Rules, including information disclosure and time-of-day of calls.
4) There must be no or minimal (one second) pause before the IVR system acknowledges that a potential respondent has answered the telephone.

Section 5.3.2 IVR Survey Introduction

Background and Questions

Section 5.3.2 states that the information disclosure requirements for IVR surveys are the same as for interviewer-administered surveys, and similarly requires that the information be provided in the survey introduction (it also addresses time of day and time-to-respond, which are addressed in the revised 5.3.1 and so are not shown here):

5. RETAINING PUBLIC CONFIDENCE

5.3. Use of Interactive Voice Response

2) When IVR is used, the same information required for interviews conducted by live interviewers (sponsor, researcher, participation is voluntary, assurance of confidentiality, etc.) must be included in the IVR survey introduction. Respondents must also be provided early in the introduction with an easy method to opt out of the survey (e.g., by pressing a specific key) so that the call is terminated gracefully and no more calls are made to that number.

For reference, information disclosure requirements are described in Section 2 *Questionnaire Design*:

2. QUESTIONNAIRE DESIGN

2.1. Standards

2) The following are required elements of all Government of Canada telephone survey questionnaire introductions:

   a) follow the *Official Languages Act* and Policies; in particular, introductions must include an active offer of both official languages (English and French); the language that is

used first will depend on the province in which the respondent resides (e.g., in Quebec "Bonjour/Hello"; in the Rest of Canada "Hello/Bonjour");

b) identify the Government of Canada or the department/agency sponsoring the survey;

c) inform respondents of the general subject and purpose of the study. The subject and purpose of the study may be expressed in very general terms so long as these terms cannot be construed as an attempt to misinform respondents. When the researcher makes a convincing case that the subject and purpose of the survey is such that stating it at the outset will affect respondents' willingness to participate, stating them may be deferred to the conclusion of the study;

d) inform respondents of the expected length of the interview;

e) identify the researcher and interviewer (a pseudonym may be used for the interviewer as long as the individual remains identifiable by management internally for quality control purposes);

f) inform respondents that their participation in the study is voluntary and completely confidential*;*

g) inform respondents that their responses remain anonymous. In the exceptional cases where the research objectives require that respondent identity be revealed, the informed consent of the respondent must be obtained;

h) inform respondents that the survey is registered with the Research Registration System maintained by the MRIA and provide information sufficient for respondents to access the System's Research Verification Service (project registration number and service's website URL, e-mail or phone number), if requested.

Examples of exceptions: The sponsor would not be mentioned until the end of the survey in particular cases (e.g., advertising post-tests); other languages would be mentioned when the survey is available in non-official languages; and information about the availability of reports could be provided to specialized respondents when the incentive to participate outweighs the additional length of the introduction.

Note: If requested by the respondent, interviewers need to be prepared to repeat the instructions. Interviewers must have information readily available to inform respondents about how respondents were chosen, how privacy is protected, and where the survey results can be obtained.

Questions Addressed by the Panel

PORD has found that IVR surveys are less than 20 minutes, and typically shorter than interviewer-administered surveys. In this context, PORD asked that the Panel comment on whether the required elements for telephone survey introductions should be revised or shortened for IVR surveys, and to comment on the possibility of moving some of the information disclosures to the end of the survey, as follows:

> ***Introduction for an IVR survey:*** *Hello/bonjour, this is (survey company) calling on behalf of the Government of Canada/government department. Pour continuez en francais appuyez sur le (number). We are conducting a (number) minute research survey on SUBJECT. Your participation is voluntary and completely confidential and anonymous. The name, telephone and web address of whom to contact for additional information about this research project are available at the end of the survey. To continue press (number). To end this call now hang up or press (number).*

(At the end of the call) *If you would like further information on this study please contact (name and phone number) or go to (website). To repeat this information press (number).*

Further information could include:

- *that the survey is registered with the Research Registration System maintained by the MRIA*
- *how to access the System's Research Verification Service (project registration number and service's website URL, e-mail or phone number)*
- *how respondents were chosen, how privacy is protected, and where the survey results can be obtained.*

Advisory Panel Response

Most Panelists said the required information in the survey introduction should be the same for IVR surveys as for other surveys.

Many supported the principle that *if* information disclosure requirements are allowed to be different for IVR surveys – either in terms of what is disclosed or where the disclosure occurs – the same allowances should be available for other surveys. Shortening the survey introduction using either method would be done to help improve response rate, and presumably any type of survey could benefit, not just IVR surveys. A few suggested additional conditions:

- Any modifications in information disclosure must be implemented in a comparable way for all surveys. That is, what information is provided and the way it is provided should be consistent for all respondents taking GC POR surveys, regardless of how the survey is administered.
- If modifications are made specifically for short IVR surveys, then the same modifications should be available for short interviewer-administered surveys.

In the context of support for retaining the existing information disclosure requirements, there was no consensus on where in the questionnaire the information should be made available. Views included:

- Stay with the existing placement in the survey introduction.
- Allow some information to be made available at the end of the survey questionnaire, as in the example provided by PORD.
- In the survey introduction, make it an option to get certain types of information. The reason for not putting this at the end was because some respondents might not agree to participate in the survey if the information is not available upfront. One comment on this approach was that providing a toll-free number or website for this purpose is logistically complex, because the respondent would need to be recontacted to see if they are willing to participate. It would be better to have the IVR system provide the information upon request.

IVR Survey Duration

Background and Questions

The standard for survey duration for telephone states surveys must be completed in 20 minutes, and strongly encourages a duration of 15 minutes or less.

2. QUESTIONNAIRE DESIGN

2.1. Standards

1) Survey questionnaires must be designed:

c) to be completed in a maximum duration of 20 minutes. Exceptions could include projects with specialized audiences and those with pre-arranged interviews when the respondent is aware the survey will take longer than 20 minutes. **Average questionnaire durations of 15 minutes or less are strongly encouraged in order to minimize respondent burden.**

The literature review commissioned by PORD suggests that IVR surveys are better suited to shorter surveys:

> One of the main drawbacks of using the IVR methodology is the need for a short, simply constructed questionnaire. This is one of the reasons this methodology is well-suited to election polling and measuring voter intention. One or two clear and concise questions can be asked when using IVR and these questions will have simple (and few) response options from which to select. For example, "press 1 if you know who you intend to vote for on election day and 2 if you do not".
>
> For a research project that intends to explore several topics with respondents and/or topics in a more in-depth manner (e.g., to uncover reasons for voter intentions or to gain insights on salient election issues with voters), an IVR survey would not be the appropriate methodology. (p.18)

Questions Addressed by the Panel

The Panel considered whether the standard for survey duration should be modified for IVR surveys.

Advisory Panel Response

The Panel agreed that IVR surveys should be kept short, and recommended adding a guideline to Standard 2.1.1. in the Telephone Standards. It was noted that IVR surveys are typically seven minutes or less in duration, and usually less than five minutes. A revised version of 2.1.1 in the Telephone Standards could be:

2. QUESTIONNAIRE DESIGN

2.1. Standards

1) Survey questionnaires must be designed:

c) to be completed in a maximum duration of 20 minutes. Exceptions could include projects with specialized audiences and those with pre-arranged interviews when the respondent is aware the survey will take longer than 20 minutes. Average questionnaire durations of 15 minutes or less are strongly encouraged in order to minimize respondent burden. For IVR surveys, durations of X minutes or less are strongly encouraged.

For the value of "X minutes", the majority of Panelists suggested five minutes as a guideline, while several suggested seven minutes. Among the latter, the comment was that five minutes is a good target for the main questionnaire, but the required information in the survey introduction may necessitate a guideline of seven minutes.

Call-back Standard for IVR surveys

Background and Questions

The call-back requirements in Section 7.2 *Call-backs* do not make any distinction between interviewer-administered surveys and IVR surveys.

> **7.2. Call-backs**
>
> 1) There will be a minimum of eight call-backs made before retiring a telephone number and substituting it with another number. The call backs must be made at varying days and times over a minimum seven-day period. An exception could be made when the field period is shorter as a result of the need to assess recall of particular events or activities.

In the literature review commissioned by PORD, under *Drawbacks of using IVR*, it states:

> In addition to limits on the length and complexity of the survey questionnaire, the quality of the sample can be questionable. This, however, is not unique to IVR. All survey research requires complete sampling frames (little to no coverage error), sound sampling strategies (simple random sampling or stratified random sampling), and appropriate sample control (an adequate number of call-backs that vary by time of day/day of the week to maximize the response rate). Since speed is one of the key advantages of using IVR, sample control measures are not as stringent (there is no time to call back a number in the sample multiple times). The survey sample will include whoever could be reached on the night of the data collection, which introduces the possibility of non-response bias. While survey weights will be applied to the survey sample post-data collection to ensure it reflects the demographic profile of target population, this will not address attitudinal differences that might exist between survey respondents and non-respondents. (p. 18)

The literature review implies that it may often be the case that few, if any, call-backs are made in an IVR survey. It connects this to using IVR for its speed advantage, and note that Standard 7.2.1 includes an exemption from the eight call-back requirement when speed of fieldwork is important: *"An exception could be made when the field period is shorter as a result of the need to assess recall of particular events or activities."*

Questions Addressed by the Panel

The Panel considered whether there should be any changes to Section 7.2 *Call-backs* specific to IVR surveys.

Advisory Panel Response

There were different views on whether and how to give call-back requirements specific to IVR surveys:

- *Exempt IVR surveys from the call-back requirements for interviewer-administered surveys:* The response rate for IVR surveys is so low that there is probably little reduction in possible non-response bias by requiring the same minimum number of call-backs as for interviewer-administered surveys. The number of call-backs, if any, should be based on the needs for each IVR project.
- *IVR surveys should have the same call-back requirements as for interviewer-administered surveys:* If it is judged that an interviewer-administered probability survey needs to have a minimum of eight call-backs (or five, as recommended by the Panel), then an IVR needs to have the same minimum number of call-backs to meet the requirements for a "good" probability survey. Under this view, 7.2.1 should be left as is.
- *IVR surveys should have the same call-back requirements as for interviewer-administered surveys – but with the understanding that in practice IVR surveys will usually have fewer call-backs:* This is basically the same as the previous position, but with the observation that probably most IVR surveys are done in a very short period of time. Because of that, the exception noted in 7.2.1 would come into play. Although, under this view, the last sentence of 7.2.1 should be modified to reflect the fact that there are other circumstances causing short fieldwork periods than just the specific use case referred to. The modification would be: *An exception could be made when the field period is shorter ~~as a result of the need to assess recall of particular events or activities~~*. Under this view, any exception from the call-back rule for an IVR survey would be triggered by its design (e.g. very short field period), not by the fact of it being an IVR survey. This means that an IVR survey with a field duration comparable to a typical interviewer-administered probability survey would have to meet the same minimum call-back requirements – although this may not be a common field duration for IVR surveys.

Multi-Mode Surveys

For reference, the current standards referring to multi-mode surveys are:

1. PROPOSAL DOCUMENTATION

1.2.4. Description of Data Collection

7) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method.

4. SAMPLING PROCEDURES

4.5. Multi-Mode Surveys

Multi-mode surveys are ones where different methods of questionnaire administration are used. They will often involve a combination of online and telephone methods, although there are other possibilities (e.g., in-person, mail, fax).

When a survey is conducted using multiple modes of questionnaire administration:

1) The reasons for using a multi-mode rather than a single-mode method must be stated, both in the research proposal and the survey report.
2) When the plan is to combine data collected via different modes in the data analyses, then steps must be taken to ensure as much comparability as possible across the different survey modes in terms of question wording and presentation of response options.
3) Steps must be taken to ensure avoidance of duplicate respondents in different modes. The steps taken, and the results, must be documented.

14 (online)/15 (telephone). MANDATORY SURVEY REPORT REQUIREMENTS

14.5/15.5 Data Collection

2) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method.

14.6/15.6 Quality Controls

3) For multi-mode surveys, detailed description of any data quality issues arising from combining data collected via different modes/instruments.

The Panel was asked to comment on possible revisions to the Standards in the following areas:

- Clarify what needs to be stated in the proposal.
- Incorporate high quality benchmark questions in order facilitate detection of mode bias.
- Encourage use of similar modes of administration – i.e. both self-administered or both interviewer-administered – in order to reduce mode effects.
- Pre-testing requirements for multi-mode surveys
- Clarify the reporting requirements in terms of (a) any adjustments made to mitigate mode biases, and (b) the decision to combine versus not combine the data across modes.
- Clarify calculation of outcome rates for multi-mode.

Proposal Documentation for Multi-Mode Surveys

Background and Questions

In Section 1, *Proposal Documentation*, the requirements specific to multi-mode surveys are:

1. PROPOSAL DOCUMENTATION

1.2.4. Description of Data Collection

7) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method.

Other components of Section 1 will ensure that the methodology employed for each mode will be described even though they don't explicitly refer to multi-mode survey designs (e.g. Section 1.2.2 *Sample/Sampling Details*, and other elements in Section 1.2.4).

The primary concern associated with multi-mode surveys is the potential for mode bias – that is, getting different response distributions for the same question area due specifically to characteristics of the mode. For example, it has been found that social desirability effects tend to be stronger in interviewer-administered surveys (e.g. telephone) than in self-completion surveys (e.g. online). Differences across mode in how questions and response options are presented could potentially cause different response distributions. Mode bias can vary by question – e.g. some questions may be more prone to social desirability effects than others, and some questions may be more similar in design across modes than other questions. These types of mode effects pose challenges for combining or comparing data across modes.

There can also be differences in response due to different types of people using the different modes, and indeed improving population coverage is a reason to consider doing multi-mode surveys in some circumstances. That said, this can also make it more challenging to detect mode bias.

The issue is whether and how the proposal documentation requirements need to be elaborated to make it more clear in the proposal that the issue of potential mode bias is recognized and that steps will be taken to address this. The existing requirement only indirectly refers to dealing with the potential for mode bias. Note that Section 4, *Sampling*, is more explicit, but perhaps more can be done in Section 1 of the Standards.

Questions Addressed by the Panel

The Panel was asked to comment on the following potential revisions to *Proposal Documentation* standards:

1. PROPOSAL DOCUMENTATION

1.2.4. Description of Data Collection

7) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method, and a rationale for the specific modes that will be used. Describe the steps that will be taken (a) to reduce the likelihood of mode biases, and (b) to facilitate detection of any mode biases.

1.2.7. Data Analysis

1) Briefly describe how the data will be analyzed in order to address the objectives/research questions, including any special analyses (e.g., segmentation)
   a) For multi-mode surveys, state whether the plan is to report results based on combining the data across modes, or to report results separately by mode.

Advisory Panel Response

Comments on Proposed Revision to Section 1.2.4

There are three basic scenarios where respondents use different modes:

1) Different respondents use different modes to respond to a questionnaire – e.g. some by telephone, and some online
2) In a multi-stage research design, different modes are used at different stages – e.g. telephone recruit to complete a survey online
3) Mode switching within a questionnaire – e.g. the first part of an interview is by telephone and then the respondent is asked to complete the questionnaire online

The Standards use the term "multi-mode" to refer to scenario #1. However, some researchers describe this as "mixed-mode" and use "multi-mode" for scenario #2. The issue of mode bias arises with scenario #1. Given the ambiguity of terminology, the Standards should include a definition of "multi-mode" to ensure everyone interprets it the same way.

The Panel agreed with the proposed revision to 1.2.4, which adds a requirement to give a rationale for the modes that will be used, and to describe steps that will be taken to reduce the likelihood of mode biases and to facilitate detection of any mode biases:

1. PROPOSAL DOCUMENTATION

1.2.4. Description of Data Collection

7) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method, and a rationale for the specific modes that will be used. Describe the steps that will be taken (a) to reduce the likelihood of mode biases, and (b) to facilitate detection of any mode biases.

Comments on Proposed Revision to 1.2.7

Most Panelists agreed with the suggested addition of a requirement to 1.2.7 to state whether the plan is to combine data across modes or to report the results separately by mode:

1. PROPOSAL DOCUMENTATION

1.2.7. Data Analysis

1) Briefly describe how the data will be analyzed in order to address the objectives/research questions, including any special analyses (e.g., segmentation)
   a) For multi-mode surveys, state whether the plan is to report results based on combining the data across modes, or to report results separately by mode.

A few noted that plans can change once the data have been collected and examined for mode effects. The suggested wording above is acceptable as long as it is understood that plans can change. However, one Panelist suggested the following revision to make this more explicit:

a) For multi-mode surveys, state that analysis will be conducted to determine the feasibility of combining the data across modes and that the conclusions of this analysis will be summarized in the report.

Sampling Procedures and *Questionnaire Design* for Multi-Mode Surveys

Background and Questions

The existing standard for multi-mode surveys in *Sampling Procedures* is:

4. SAMPLING PROCEDURES

4.5. Multi-Mode Surveys

Multi-mode surveys are ones where different methods of questionnaire administration are used. They will often involve a combination of online and telephone methods, although there are other possibilities (e.g., in-person, mail, fax).

When a survey is conducted using multiple modes of questionnaire administration:

1) The reasons for using a multi-mode rather than a single-mode method must be stated, both in the research proposal and the survey report.
2) When the plan is to combine data collected via different modes in the data analyses, then steps must be taken to ensure as much comparability as possible across the different survey modes in terms of question wording and presentation of response options.
3) Steps must be taken to ensure avoidance of duplicate respondents in different modes. The steps taken, and the results, must be documented.

There is no current standard for *Questionnaire Design* specific to multi-mode surveys.

Questions Addressed by the Panel

For purposes of Panel discussion, changes were suggested in order (a) to distinguish between sampling-related topics and questionnaire-related topics, (b) to increase the prominence of the value of using similar modes of survey administration, (c) to clarify that one needs to be concerned about mode biases when comparing results by mode as well as when combining data across modes, and (d) to highlight the value that benchmark questions can have for enabling detection of mode biases.

The proposed revisions to Section 4 *Sampling Procedures* were:

4. SAMPLING PROCEDURES

4.5. Multi-Mode Surveys

Multi-mode surveys are ones where different methods of questionnaire administration are used (e.g. some combination of telephone, online, in-person, or mail).

When a survey is conducted using multiple modes of questionnaire administration:

1) The reasons for using a multi-mode rather than a single-mode method must be stated, both in the research proposal and the survey report.
2) The rationale for the specific modes used must be stated, both in the research proposal and the survey report.
    a) The risk of mode biases can be lower if the modes of administration are similar – i.e. both interviewer-administered (e.g. telephone and in-person) or both self-administered (e.g. online and mail).
3) Steps must be taken to ensure avoidance of duplicate respondents in different modes. The steps taken, and the results, must be documented.

The proposed revisions to Section 2 *Questionnaire Design* were as follows -- this would be a separately numbered item in Section 2, shown here as "x":

2. QUESTIONNAIRE DESIGN

2.1. Standards

x) Multi-mode surveys are ones where different methods of questionnaire administration are used (e.g., some combination of telephone, online, in-person, or mail).

When a survey is conducted using multiple modes of questionnaire administration:

a) When the plan is to combine or compare data collected via different modes in the data analyses, then steps must be taken to ensure as much comparability as possible across the different survey modes in terms of question wording and presentation of response options.

b) Include questions for which there are high quality data on the population and which therefore can be used as benchmarks for exploring the possibility of mode biases. The required demographic questions can serve this purpose (see Section 1.2.4). Also try to identify any other high quality benchmarks available – whether demographic, behavioural or attitudinal –that might correlate with key measures in the survey. The latter can help better identify any mode biases in the key survey measures.

Advisory Panel Response

Proposed Revision to Section 4.5 *Multi-Mode Surveys*

The Panel agreed with the proposed revision to Section 4.5, which emphasizes the value of using similar modes of administration:

4. SAMPLING PROCEDURES

4.5. Multi-Mode Surveys

Multi-mode surveys are ones where different methods of questionnaire administration are used (e.g. some combination of telephone, online, in-person, or mail).

When a survey is conducted using multiple modes of questionnaire administration:

1) The reasons for using a multi-mode rather than a single-mode method must be stated, both in the research proposal and the survey report.
2) The rationale for the specific modes used must be stated, both in the research proposal and the survey report.
    a) The risk of mode biases can be lower if the modes of administration are similar – i.e. both interviewer-administered (e.g. telephone and in-person) or both self-administered (e.g. online and mail).
3) Steps must be taken to ensure avoidance of duplicate respondents in different modes. The steps taken, and the results, must be documented.

Proposed Addition of a Multi-Mode Standard to Section 2.1 *Questionnaire Design*

The Panel agreed with the suggested addition to Section 2.1, with one modification. The point was made that while use of behavioural and attitudinal benchmarks can be useful in detecting mode bias, this analysis needs to control for any differences in demographic variables. The standard was revised to include this cautionary note:

2. QUESTIONNAIRE DESIGN

2.1. Standards

x) Multi-mode surveys are ones where different methods of questionnaire administration are used (e.g., some combination of telephone, online, in-person, or mail).

When a survey is conducted using multiple modes of questionnaire administration:

a) When the plan is to combine or compare data collected via different modes in the data analyses, then steps must be taken to ensure as much comparability as possible across the different survey modes in terms of question wording and presentation of response options.

b) Include questions for which there are high quality data on the population and which therefore can be used as benchmarks for exploring the possibility of mode biases. The required demographic questions can serve this purpose (see Section 1.2.4). Also try to identify any other high quality benchmarks available – whether demographic, behavioural or attitudinal –that might correlate with key measures in the survey. After controlling for any differences by mode on key demographics, the latter can help better identify any mode biases in the key survey measures.

Two other important issues came up in the Panel discussion, although these probably do not have direct implications for standards:

- The "Don't know" response option (and related forms such as "Prefer not to answer") poses a challenge when one mode is interviewer-administered (e.g. telephone) and the other mode is self-completion (e.g. online). Typically, "Don't know" is not explicitly presented in an interviewer-administered survey. In an online self-completion mode, this response option can be presented, although it does not have to be. Explicitly presenting the "Don't know" option in one mode but not the other likely will affect the response distribution. Not presenting the "Don't know" option in the online mode is not exactly comparable to the telephone approach, because the respondent is either forced to choose one of the substantive response options, or they may think they have to choose one of those options (i.e. if they do not realize they can proceed without answering the question). These issues support the Panel's previous addition to 4.5, that *the risk of mode biases can be lower if the modes of administration are similar – i.e. both interviewer-administered (e.g. telephone and in-person) or both self-administered (e.g. online and mail).*
- If a mode effect exists, the standards require disclosure and discussion of this in the report, but there may not be any way to mitigate this effect in the results.

Pre-Testing for Multi-Mode Surveys

Background and Questions

The current Section 3 standard for pre-testing does not make any specific references to multi-mode surveys.

For reference, Section 3 *Pre-Testing* (online version) is:

3. PRE-TESTING

3.1. Standards

1) Pre-testing of all components of a new or revised survey questionnaire that may influence data quality and respondent behaviour is required. This includes the online appearance and functionality of the questionnaire.
2) The client must be given the opportunity to test and approve the online survey prior to launch.
3) Pre-testing must include probing that invites participants recruited for this purpose to provide input about their comprehension of and reaction to the questions. For example, a short series of questions could be included at the end of the pre-test survey. Researchers and clients must agree in advance as to whether probing will take place during or after administering the survey. If requested by the client a cognitive pre-test must be conducted.
4) To help ensure questionnaire effectiveness with subgroups where there is reason for concern (e.g., due to language, age, level of education, etc.), the socio- demographic characteristics of the targeted participants must be approved by the client before recruiting begins.
5) A minimum of 10 pre-test surveys are to be completed in each language in which the final survey will be fielded. An exception could be projects with small survey populations, such as a client-satisfaction survey of a small client base. In such cases the researcher must, in consultation with the client, take steps to ensure that the smaller number of pre-tests are sufficient to guarantee questionnaire quality. For example, a cognitive pre-test may be warranted.
6) Pre-test completions shall not be included in the final dataset. An exception could be projects with:
   a) hard-to-reach target groups, or
   b) when no changes are made to the questionnaire.
7) Documentation of the pre-test(s) must be provided to the client before the questionnaire is finalized. The documentation must include (at minimum):
   a) a description of the pre-test approach and number of interviews completed;
   b) findings and any resulting modifications;
   c) average survey completion time;
   d) a statement of whether or not pre-test cases will be retained in the final data set.

   The final research report must include this same information.

Questions Addressed by the Panel

The Panel was asked to consider whether for a multi-mode survey there should be a requirement for a minimum number of pre-test interviews in English and French for each mode.

Advisory Panel Response

Based on the current 3.1 standards, a full pre-test in each mode would mean doing 10 English and 10 French surveys in each mode.

One Panelist recommended a standard of doing a full pre-test in each mode, and another recommended doing a full pre-test unless the questionnaires are identical across modes – which may not commonly be the case.

Other Panelists said it depends on how similar the different modes are to each other:

- If both modes are interviewer-administered (e.g. telephone and in-person) or both are self-completion (e.g. online and mail), then the existing 10 minimum by language would apply, providing within a language the 10 interviews are a mix of the two modes.

  It was noted that there can be reasons to do more than a total of 10 per language depending on the circumstances. For example, when both are interviewer-administered (e.g. telephone and in-person), there may be logistical issues specific to the in-person interviewing that would benefit from additional pre-tests. However, this would be determined on a case by case basis by the researcher as to whether additional pre-tests are needed.
- If one mode is interviewer-administered and another mode is self-completion, then there should be a full pre-test in each mode.

Outcome Rates for Multi-Mode Surveys

Background and Questions

Currently in Section 8 *Outcome Rates* there is no standard for how to calculate outcome rates for a multi-mode survey.

Questions Addressed by the Panel

The Panel was asked to consider adding a standard to Section 8 Outcome Rates that addresses how to approach determining outcome rates for a multi-mode study. For discussion purposes, the following added standard to Section 8 was suggested:

8. OUTCOME RATES

8.x. Multi-mode surveys

1) The general principles for calculating the response rate for mixed-mode surveys are (a) calculate a response rate for each mode or stage of sampling, (b) use the same general response rate formula for each mode or stage of sampling, and (c) calculate a single

aggregated response rate. The specific details of how this is done will be adapted to the particular multi-mode survey design.

Advisory Panel Response

There are two problems with the suggested addition to Section 8 *Outcome Rates*:

- It uses the term "response rate", but the Standards currently, and appropriately, use "outcome rate." This is because an outcome rate can be either a "response rate" in the case of probability surveys, or a "participation rate" in the case of non-probability surveys.
- When the outcome rate for one mode is a response rate and the outcome rate for another mode is a participation rate, it is not appropriate to aggregate these two rates. Also, if the data from different modes are not being combined in the data analysis, then again the outcome rates should not be aggregated.

The following addresses these issues:

> **8. OUTCOME RATES**
>
> **8.x. Multi-mode surveys**
>
> 1) The general principles for calculating the mandatory outcome rates for mixed-mode surveys are (a) calculate the mandatory outcome rate for each mode or stage of sampling, and (b) if appropriate, calculate a single aggregated outcome rate. The specific details of how this is done will be adapted to the particular multi-mode design. An acceptable approach could involve presenting separate outcome measures for each mode or stage of sampling.

Also discussed were research designs where no outcome rate can be calculated, and therefore Section 8 of the Standards cannot be applied. Two such scenarios are:

- In a Panel survey, a respondent is sent multiple survey invitations at the same time, and the respondent chooses which, if any, of the surveys to complete.
- A sample router is used. A sample router may be used by an online sample provider, and it is software that screens potential respondents and then uses some procedure or rule to assign them to one of multiple waiting surveys.

In neither case is it clear how to calculate a meaningful outcome rate.

The use of sample routers was addressed in a 2016 AAPOR report, *Evaluating Quality in Today's Complex Environment*, and it noted the following:

> More recently, online sample providers have moved aggressively to the use of sample routers—software that screens potential respondents and assigns them to one of many waiting surveys… This can improve the efficiency of fieldwork, but there is little published research on the effect of routers on bias. One concern is the degree to which sample composition is affected by other surveys that are active in the router at the same time.

The Panel recommended that research designs which do not allow calculation of either of the mandatory outcome rates should not be permitted for GC POR surveys. That is, for GC POR surveys, the sample of respondents to be contacted should be well defined, and each respondent should be invited to complete a single survey questionnaire.

Mandatory Survey Report Requirements for Multi-Mode Surveys

Background and Questions

The existing standards in the *Mandatory Survey Report Requirements* referring to multi-mode surveys are:

14 (online)/15 (telephone). MANDATORY SURVEY REPORT REQUIREMENTS

14.5/15.5 Data Collection

2) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method.

14.6/15.6 Quality Controls

3) For multi-mode surveys, detailed description of any data quality issues arising from combining data collected via different modes/instruments.

Section 14.5/15.5.2 should be updated based on the Panel discussion of Section 1.2.4 in *Proposal Documentation*, where this same language was used. Based on the Panel's recommendation for Section 1.2.4, Section 14.5/15.5 should be revised as follows:

14 (online)/15 (telephone). MANDATORY SURVEY REPORT REQUIREMENTS

14.5/15.5 Data Collection

2) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method, and a rationale for the specific modes that were used. Describe the steps taken (a) to reduce the likelihood of mode biases, and (b) to facilitate detection of any mode biases.

Questions Addressed by the Panel

The Panel considered the following possible expansion of Section 14.6/15.6 *Quality Controls*, with the intent (a) to ensure there is clarity as to decisions made about combining or not combining data across modes, and (b) to require descriptions of any adjustments made to the data to mitigate mode effects:

14.6/15.6 Quality Controls

3) For multi-mode surveys, detailed description of any data quality issues arising from combining data collected via different modes/instruments. The discussion must also include:
   a) Rationale for decisions about combining or not combining data from different modes in reporting the results.
   b) Description of any steps taken to mitigate mode effects in the survey results.

Advisory Panel Response

The Panel agreed with the suggested revision to Section 14.6/15/6, although there was a minor disagreement over the wording of part "b". Some agreed with the wording proposed, but another view was to use the following wording: *Description of any adjustments made to the data to mitigate mode effects in the survey results*. The argument in favour of "steps" was that there might be sequential operations taken on the data to mitigate mode effects (e.g. data cleaning, weighting).

Incentives in Surveys of Children, Young People or Vulnerable Respondents

Background and Questions

Section 6 *Data Collection from Children, Young People or Vulnerable Respondents* does not make any reference to whether or how incentives are used for this survey population.

Section 7.5 (telephone)/7.6 (online) does not refer to this population either:

> **7.5/6. Incentives/Honoraria**
>
> 1) The details of any incentives/honoraria to be used for an telephone/online survey must be provided in both the proposal and report documentation, including:
>   a) the type of incentive/honoraria (e.g., monetary, non-monetary);
>   b) the nature of the incentive (e.g., for monetary, prize, points, donations, direct payments);
>   c) the estimated dollar value of the incentives to be disbursed.

GC's *Qualitative Research* standards contain one standard pertaining to incentives for this target population:

> **5. Participant Recruiting**
>
> **5.4. Incentives**
>
> 7) Ethical considerations should be taken into account especially when providing incentives to youth, or vulnerable groups (e.g. payment is not coercive, or exposes young or vulnerable persons to a risk that they would otherwise have avoided).

Note that the Standards define children and young people as follows:

- Children: Under the age of 13
- Young people: 13-15

People 16 and over do not require the consent of a parent or responsible adult (guardian, etc.) before being invited to participate in a survey.

Questions Addressed by the Panel

The Panel considered whether or not there should be guidance on incentives/honoraria for children, young people or vulnerable individuals who participate in a telephone or online survey. For discussion purposes, the following addition to the *Incentives/Honoraria* standard was suggested.

> **7.5/6. Incentives/Honoraria**
>
> 1) The details of any incentives/honoraria to be used for an telephone/online survey must be provided in both the proposal and report documentation, including:
>   a) the type of incentive/honoraria (e.g., monetary, non-monetary);
>   b) the nature of the incentive (e.g., for monetary, prize, points, donations, direct payments);

c) the estimated dollar value of the incentives to be disbursed.

2) When survey respondents are children, young people or vulnerable individuals, and an incentive is being offered:
   a) Decide in advance who will receive the incentive – the parent or responsible adult (guardian, etc.), the respondent, or if both will receive an incentive.
   b) The parent or responsible adult must agree to the incentive, regardless of who is receiving the incentive.
   c) Ethical considerations should be taken into account when providing incentives to children, youth, or vulnerable groups (e.g. payment is not coercive, or exposes young or vulnerable persons to a risk that they would otherwise have avoided).

Advisory Panel Response

There should be guidance in the telephone and online standards on incentives/honoraria for children, young people and vulnerable respondents.

The majority of Panelists agreed with the suggested wording as is, but there were two suggested changes by a few Panelists:

- Revise 2a to make providing the incentive to the respondent the default option, and require justification if it is to be provided to the parent or responsible adult.
- Revise 2a to emphasize the role of the study requirements in making the decision. For example: *Decide in advance who will receive the incentive based on the particular requirements of the study – the parent or responsible adult (guardian, etc.), the respondent, or if both will receive an incentive.*

Privacy and Security of Data

Passive Data Collection in Online Surveys

Background and Questions

Online and mobile methodologies create possibilities for collecting various types of personal data "passively", that is without direct interaction with respondents. The issue to be considered is, in the context of surveys, what passive data collection is allowed and under what circumstances is it allowed?

The ESOMAR/GRBN *Guideline on Mobile Research*[26] states:

> Mobile applications are capable of collecting a broad range of personal data without direct interaction with data subjects. Examples include web use and browsing history, app usage statistics, loyalty card data, geolocation, social media data, data from wearables and IoT and other data generated by or obtained from mobile devices.

At least some of these capabilities can also apply to online surveys where the device used by the respondent is a computer.

The *Guideline* notes an important exception to what constitutes "personal" data involving certain device characteristics:

> While it is possible to passively detect the type of device a data subject is using, this is not personal data as long as the purpose is to optimize app performance and survey rendering.

The existing standards address relevant general principles of protection of respondent anonymity and confidentiality, and address passive data collection to some extent.

- Section 5.1.4, *Protection of Anonymity and Confidentiality*, states the general principle: *The anonymity of respondents must always be preserved unless they have given their informed and explicit consent to the contrary.*

  5.1.4. Protection of Anonymity and Confidentiality

  1) The anonymity of respondents must always be preserved unless they have given their informed and explicit consent to the contrary. If these respondents have given informed consent for data to be passed on in a form which allows them to be personally identified, the researcher must ensure that the information will be used for research purposes only, OR, if requested by the respondent, to resolve a customer complaint. The same holds true when respondents' answers are collated, with their informed consent, with pre-existing data that allows such identification (e.g., with an administrative data base). Moreover, such personally identifiable information must not be used for any purpose unrelated to the current study, such as direct marketing, list-building, credit rating, fund-raising, or any marketing activities directed at those individual respondents.

  Section 5.1.4 establishes that any personal data collected passively must be handled in a way that protects the anonymity and confidentiality of the survey respondents.

[26] ESOMAR and GRBN. (2017). *Global Guideline on Mobile Research*

- Section 5.3, *Privacy Issues Specific to Online Survey research*, states a requirement to have an accessible policy statement *"concerning the use of cookies, log files and, if applicable, software."* This ensures respondents are informed of certain types – but not all types – of passive data collection.

> **5.3 Privacy Issues Specific to Online Survey Research (Online)**
>
> 1) Researchers must have a readily accessible policy statement concerning the use of cookies, log files and, if applicable, software. This statement may be either included in their privacy policy or it may appear in a separate document. Software must not be installed on respondents' computers without their knowledge or consent. In addition, respondents must be able to remove the researcher's software easily from their machines (e.g., for Windows users, the software must appear in the Add/Remove Programs folder in their Control Panel).
> 2) Any links to data protection, privacy policy or cookie policy statements must be given at the start of the questionnaire.

- Section 7.2, *Data Collection and Recruitment Techniques* (Online Standards), addresses passive data collection with its reference to forbidding "surreptitious" or "unsolicited" data collection – terms which appear to refer to passive data collection generally. The principle is that passive data collection can only be done with the respondent's awareness and presumably consent (although the standard doesn't say "consent"):

> **7.2. Data Collection and Recruitment Techniques**
>
> 1) Researchers must not make use of surreptitious, misleading or unsolicited data collection or recruitment techniques – including using spambots, spiders, sniffers or other 'agents' that collect personal information without the respondent's explicit awareness.

It may be that the above standards are sufficient to address issues associated with passive data collection in surveys. Alternatively, it may be that the existing standards are not sufficiently explicit. To aid in Panel consideration of this topic, standard 7.2 was revised as follow:

> **Possible revision of 7.2. Data Collection and Recruitment Techniques**
>
> 1) Passive data collection refers to collecting personal information about respondents without direct interaction with them.
>
> Researchers must not make use of data collection or recruitment techniques using passive data collection methods unless survey respondents or potential respondents have first given informed consent, or unless collecting the information is legally permissible or is permissible under the Terms of Use of the website, service or application from which the data are sourced.
>
> Examples of passive data collection of personal information include, but are not limited to, web use and browsing history, app usage statistics, geolocation, personally identifiable biometric data, social media data, data from wearables and IoT (internet of things), and other data generated by and obtained from respondents' mobile devices or computers.
>
> Passive detection of the type of device a respondent is using is not personal data as long as the purpose is to optimize app performance and survey rendering.

Some things considered when formulating this possible revision to 7.2:

- The Panel is reviewing standards for online and telephone surveys, so the language emphasizes the survey context for the standard.
- The original 7.2 referred to both "data collection and recruitment techniques", so both have been retained.

- The proposed revision includes a definition of "passive data collection" of personal information.
- The ESOMAR/GRBN standards make references to exceptions where the data collection is "legally permissible" or is allowed under "Terms of Use." This has been incorporated into the proposed revision of 7.2.
  - -- ESOMAR/GRBN *Guideline on Mobile Research*, Section 3.4 on *Passive data collection*: *"When it is not possible to obtain consent (such as measuring traffic to a website), researchers must have legally permissible grounds to collect the data…"* (p. 10).
  - -- ESOMAR/GRBN *Global Guideline – Online Research*, Section 3.2.1 *Passive data*:

    > Where personal data are collected from public spaces such as websites or social media sites, consent must be obtained directly or explicitly provided for in the Terms of Use (ToU) policy of the platform. This does not apply to publication in social media that includes the author's name, which implies a diminished expectation of privacy. (p.10)

    Note: The last sentence referring to not applying to *"publication in social media that includes the author's name"* arguably does not apply to survey research. Even if there is passive collection of such information, it forms part of a data record that includes responses to a survey questionnaire. Standard 5.1.4 ensuring anonymity and confidentiality therefore takes precedence.
- Another reason for the inclusion of "legally permissible" and "Terms of Use" involves surveys done using a panel or a market research online community (MROC). Such surveys may incorporate information available on panel/community members. This would be allowed providing it is permissible under the Terms of Service of the panel or MROC.
- The ESOMAR/GRBN *Guideline on Mobile Research*, Section 3.4 *Passive data collection*, includes a subsection 3.4.1 *Biometric data*. The examples they cite:

  > For example, facial coding involves recording a data subject's face as he or she completes a survey or similar task. Eye tracking, virtual reality headsets and other wearable devices may be used in a similar way. (p.10)

  To address this, the proposed revision to 7.2 includes *"personally identifiable biometric data"* (not all biometric data is personally identifiable; for example, an eye-tracking gaze pattern superimposed on the object being looked at does not identify the individual).
- The proposed revision explicitly excludes device type detection from the informed consent requirement. It uses the language in the ESOMAR/GRBN *Guideline on Mobile Research* (the footnote on p.10).

Questions Addressed by the Panel

The Panel considered whether the current standards adequately address passive data collection of personal information done in conjunction with a survey. For discussion purposes, the Panel considered the proposed revision to 7.2 *Data Collection and Recruitment Techniques*.

Advisory Panel Response

The Panel endorsed the proposed revision to 7.2 as an improvement over the current version, with one modification. The modification is to the last paragraph stating that device type is not considered to be personal information. In order to optimize survey rendering of an online survey, it

is also useful to have information on browser characteristics. The last paragraph should be revised as follows:

> **Original proposal**
>
> Passive detection of the type of device a respondent is using is not personal data as long as the purpose is to optimize app performance and survey rendering.
>
> **Panel recommendation**
>
> Passive detection of the browser characteristics and settings including the type of device a respondent is using is not personal data as long as the purpose is to optimize app performance and survey rendering.

One Panelist suggested checking the proposed revision to 7.2 with the Office of the Privacy Commissioner of Canada to see if there are any other best practices that should be incorporated into the standard.

Although this topic pertained specifically to online surveys, there was some discussion of whether a standard on passive data collection needs to be added to the Telephone Standards. Overall, it is not clear that there are any legally permissible forms of passive data collection in telephone surveys, and so it appears that it is not necessary to add anything about passive data collection to the Telephone Standards.

Note that an audio recording of a telephone survey call can be considered a form of data collection. However, Section 5.1.4 #4 of the Telephone Standards states that the respondent must be informed of the recording, so "passive" audio recording is not permitted:

> **5.1.4 Protection of Anonymity and Confidentiality**
>
> 4) When recording or observation techniques are used, researchers must advise the respondent at the beginning of the interview of this fact. Respondents' confidentiality must not be jeopardized by the use of such methods.

Photographs and Recordings

Background and Questions

The Online and Telephone Survey Standards do not currently have any standards pertaining specifically to respondent photographs, videos or audio recordings.

The ESOMAR/GRBN *Guideline on Mobile Research* Section 3.4.2 states the following about *Photographs and recordings:*

> Photographs, video and audio recordings are considered to be personal data and therefore must be gathered, processed and stored as such. They can only be shared with a client if the data subject gives his or her prior consent with knowledge of the specific purpose for which it will be used. When potentially identifying information has been removed (such as through pixelisation or voice modification technology) so that it is no longer considered personal data it can be shared with a client provided the client agrees to make no attempt to identify the individual.
>
> Researchers must not instruct data subjects (or those that may be acting as data collectors) to engage in surveillance of individuals or public places. Data subjects should be

given specific limited tasks (e.g. capturing interactions with friends with their consent, or images of objects or displays) that do not involve monitoring a particular area where personal data would be captured without the consent of the individuals present. When recorded observation of a location is undertaken, clear and legible signs indicating that the area is under observation along with the contact details for the researcher or research organization performing the research should be posted and images of individuals must be pixelated or deleted as soon as possible. Cameras should be situated so that they monitor only the areas intended for observation. (pp. 10-11)

The following is a slightly revised version of the ESOMAR/GRBN guideline that emphasizes the survey context (as the Panel is only considering revisions to the Telephone and Online Survey Standards):

> **Revised to emphasize survey context:** Photographs, video and audio recordings from survey respondents are considered to be personal data and therefore must be gathered, processed and stored as such. They can only be shared with a client if the respondent gives his or her prior consent with knowledge of the specific purpose for which it will be used. When potentially identifying information has been removed (such as through pixelisation or voice modification technology) so that it is no longer considered personal data it can be shared with a client provided the client agrees to make no attempt to identify the individual.
>
> Researchers must not instruct survey respondents to engage in surveillance of individuals or public places. Respondents should be given specific limited tasks (e.g. capturing interactions with friends with their consent, or images of objects or displays) that do not involve monitoring a particular area where personal data would be captured without the consent of the individuals present. When recorded observation of a location is undertaken, clear and legible signs indicating that the area is under observation along with the contact details for the researcher or research organization performing the research should be posted and images of individuals must be pixelated or deleted as soon as possible. Cameras should be situated so that they monitor only the areas intended for observation. (pp. 10-11)

Questions Addressed by the Panel

The Panel considered whether or not the modified ESOMAR/GRBN *Guideline on Mobile Research* Section 3.4.2 should be added to the Online and Telephone Survey Standards.

Advisory Panel Response

The first paragraph of the modified ESOMAR/GRBN guidance should be incorporated into the updated standards. One Panelist also noted that while at present this mainly applies to online surveys, it can apply to telephone surveys where the respondent is using a smart phone. For example, the respondent could be asked to take a picture of something and send it to the survey company (e.g. attached to a SMS message sent to the calling phone number). Therefore, it should be included in both the Online and Telephone Standards.

With regard to the second paragraph of the ESOMAR/GRBN guidance, there was discussion of whether this is a sufficiently common scenario to justify inclusion in the Standards. The first sentence forbids instructing respondents to engage in "surveillance of individuals or public places." The issue some had with this is that asking for "surveillance" would be very unlikely in GC POR, and so it may not be worth addressing in the Standards.

The second sentence in the second paragraph addresses a more realistic scenario of instructing a respondent to take an image or recording of something relevant to the survey topic. Particularly because of the prevalence of smartphones, this may increasingly be one of the tools used by survey researchers for certain types of survey topics (e.g. a visitor survey). A simplified version of the ESOMAR/GRBN text is worth including in the updated standards. The following is a possible wording:

> **Revised second paragraph:** If respondents are asked to make recordings as part of their survey input, they should be given specific limited tasks (e.g. capturing interactions with friends with their consent, or images of objects or displays) that do not involve monitoring a particular area where personal data would be captured without the consent of the individuals present.

Telephone Surveys – Sensitivity to Setting

Background and Questions

PORD requested the Panel consider whether there needs to be an addition to the Telephone Standards related to sensitivity to the respondent's setting. In this regard, there are two relevant guidelines in the ESOMAR/GRBN *Guideline on Mobile Research* (p. 7):

> **3.1.1 Safety**
>
> When calling mobile phones researchers may sometimes contact potential data subjects who are engaged in an activity or in a setting not normally encountered in fixed-line calling. This might include driving a vehicle, operating machinery or walking in a public space. The researcher should confirm whether the individual is in a situation where it is legal, safe and convenient to take the call. If the researcher does not receive confirmation, then the call should be terminated while allowing the possibility of making further attempts at another time.
>
> **3.1.2 Confidentiality and sensitive data**
>
> A researcher might contact a potential data subject who is engaged in an activity or situation where others may overhear the call. In this case, the researcher must consider the nature of the research content in light of the possibility that the data subject might be overheard and personal information or behaviour inadvertently disclosed or responses modified as a result of their situation. If appropriate, the call should be rescheduled to another time or location when confidentiality will not to be compromised.

Note that ESOMAR/GRBN 3.1.1 above is specific to mobile phones, but 3.1.2 could apply to either mobile or fixed-location phones.

The current Telephone Standards, in Section 5.2.1 *Avoidance of Harassment*, has a standard focused on sensitivity of the survey subject matter, but it does not directly address issues caused by the setting of the interview:

> **5.2. Avoidance of Harassment**
>
> 1) The researcher must take all reasonable steps to ensure that respondents are not in any way hindered or embarrassed by any interview, and that they are not in any way adversely affected as a result of it. Researchers must address sensitive subject matter in a way that will minimize the discomfort and apprehension of both respondents and interviewers.

Questions Addressed by the Panel

Because respondents are increasingly likely to answer calls using a mobile phone, there can be issues with them using the phone in problematic settings (e.g. driving, walking in a public space). On both mobile phones and fixed-location phones, they may be in a setting where they can be overheard.

The Panel considered whether the existing standards are sufficient, or whether there should be an additional standard or guideline for the interviewer to confirm the respondent is in a location where they are comfortable taking the call.

Advisory Panel Response

Most Panelists supported adding a guideline to determine if a telephone survey respondent is in a location where they can take the call. One Panelist suggested the following wording:

> The interviewer should confirm with a respondent that they are in a location where they are comfortable doing the interview. A survey introduction could include a question similar to the following: *Is this a safe and convenient time to conduct an interview?* This is particularly advisable where the respondent is on a cell phone. If the researcher expects that the respondent's setting may hinder their ability to participate in the survey, the researcher may take additional steps such as rescheduling the interview.

The reasons most recommended a guideline rather than a standard is that while this is a good practice, it can be problematic when a survey is at risk of being overly long. Further, it was suggested that often the benefit of asking the question may be small because:

- It is reasonable to expect that people will be responsible deciding for themselves, without being asked, whether or not they can or should take the call or do the interview.
- The existing standard 5.2*, Avoidance of Harassment*, arguably already deals with the issue of asking sensitive questions in situations where the respondent can be overheard.

One Panelist said this should be a standard – i.e. required – for all telephone surveys. The reason is that especially for respondents using a cell phone, some may be tempted to do the survey even though they are in a setting such as driving where it is not entirely safe to do so. In order to avoid potentially contributing to unsafe behaviour, it was suggested a standard is needed. Further, it was suggested that cell phone samples are not guaranteed to be 100% cell phones, so this should be required for all telephone surveys, not just the portion of the survey directed to a cell phone sample.

Data Breaches

Background and Questions

A data breach is the loss of or unauthorized access to/disclosure of personal or organizational information.

The current standards require taking steps to protect against data breaches. The objective for the Panel is to identify any revisions or additions to the standards, and/or any guidelines that should be included.

The current standards are:

13.2 (Online)/14.2 (Telephone). Protection of Data/Servers

1) Protection against illegal or unsanctioned access: Researchers must use up-to-date technologies to protect survey data collected or stored on Web sites or servers against illegal or unsanctioned access by third parties (i.e. "hacking"). The researcher must also control access to all databases on which any data relating to the survey is stored so that only individuals with the appropriate security clearance are able to access the database, either by using a password or other form of access restriction (such as biometric controls).
3) The researcher must not subcontract (including to an affiliate) any function that involves providing a subcontractor with access to any data relating to the survey unless the client first consents in writing.
5) Protection against physical damage to servers: Researchers must also put in place measures to ensure the "physical" security of data and servers.

13.3/14.3 Temporary Storage of Data on Servers

1) If the temporary storage of data collected takes place on a server that is operated by another provider, the researcher must place the provider under the obligation to take the necessary steps to ensure that the requirements described in subsection 13.2 are met. Temporary storage of the collected data on the server must be terminated at the earliest possible time.

13.6/14.5 In the Event of Any Data Breach

1) In the event of any data breach, the client must be informed immediately and provided with details about both the nature and the extent of the data breach.

The literature review commissioned by PORD (see pp. 28-30) cites the following as a framework for considering data protection:

- **Be aware of the data the organization has** – Know exactly what kind of data the organization has, where/how it is stored, as well as where/when it is collected, and who has access. When organizations have a clear understanding of the data, they can identify the type of data that would require a unique protection system and they can adopt or develop approaches to safeguard these data.
- **Be aware of the organization's vulnerabilities** – Risk and vulnerability assessments help to ensure that threats to privacy are identified and addressed. The vulnerabilities organizations should be aware of include: third-party activities involving the organization and data.
- **Limit the information collected and the length of time the information is retained** – It is not only important to know why you collect the information, but organizations should know why they are holding this information.
- **Clearly define policies and procedures about the secure destruction of information** – Improper disposal of the information can lead to data leakage.
- **Train employees** – Policies can only be effective when those responsible for implementing and abiding by them are aware of what they contain, why they exist, and the consequences of neglecting their responsibilities.
- **Maintain up-to-date software and safeguards** – Establish routine and documented steps to ensure security-related updates are applied in a timely manner, and that software no longer in use is removed from the system.

- **Implement and monitor intrusion prevention and detection systems** – Measures such as intrusion detection systems, firewalls and audit logs can help to identify and respond to privacy breaches before they escalate.
- **Encrypt laptops, USB keys and other portable media**.

PORD requested the Panel to consider whether this framework suggests any revisions to the current standards cited above.

Questions Addressed by the Panel

The Panel considered whether any revisions are needed to the current standards pertaining to the data privacy and security standards cited above.

Advisory Panel Response

The existing standards pertaining to privacy and security of data, including data breaches, are appropriate.

There were two areas where there was some discussion of the need for additional standards or guidelines: encryption of data on portable devices, and data retention limits.

Encryption of data on portable devices: There should be guidance on encryption of survey data stored on portable devices (laptops, USB drives and other portable media). The Panel distinguished between survey data files that contain personally identifiable information (PII) and those that do not. There should be a requirement to encrypt survey data on portable devices with PII, and a guideline to encrypt when there is not PII.

Limits on retention of PII. The current standards address the minimum time for which survey data must be retained (*Retention of Technical Data* – 13.1 of Online Standards, and 14.1 of Telephone Standards). However, there is no standard requiring destruction of PII after a period of time.

Limits on data retention are addressed in some other contexts:

- *MRIA:* Article 7 of the *Code of Conduct*, Part c *Use of data*, states: *PII collected and held in accordance with this Code shall be…: (c) Preserved no longer than is required for the purpose for which the information was collected or further processed.*
- *PIPEDA:* One of the PIPEDA fair information principles[27] related to limiting retention of personal information is:

  Principle 5 – Limiting Use, Disclosure, and Retention

  Unless the individual consents otherwise or it is required by law, personal information can only be used or disclosed for the purposes for which it was collected. Personal information must only be kept as long as required to serve those purposes.

[27] PIPEDA, *PIPEDA fair information principles*. Reviewed January 2018. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/p_principle/

- *GC departments/agencies:* One Panelist noted that GC departments/agencies sometimes stipulate a data retention limit.

It was suggested that the Standards should add rules about retention limits on PII, although the Panel was divided on whether this should be a standard or a guideline, and on whether it should leave the time frame for retention general in nature (as is the case with MRIA and PIPEDA), or state a specific time frame. With regard to the latter, one Panelist suggested a guideline of the following sort:

> *PII data for a project (either received from a client for purposes of data collection, or collected as part of the study) should be retained for no more than one year following the end of the contract unless there are legitimate reasons for retaining said data beyond one year.*

The rationale for stating a specific time frame is that it is easier to explain should a respondent enquire about this. The specific time frame could be something other than a year, but one year arguably gives sufficient latitude for the client and supplier to deal with any survey-related queries or issues that might arise.

Other possible additional guidelines

Some suggestions for additional guidelines included:

- Conduct annual risk and vulnerability assessments
- Train employees in data protection policies
- Clearly define policies and procedures about the secure destruction of information
- Maintain up-to-date software and safeguards by installing security-related updates in a timely manner
- Implement and monitor intrusion prevention and detection systems.

Cloud Storage

Background and Questions

The current standards require that survey data be stored in Canada:

13.2/14.2 Protection of Data/Servers

2) Protection against legally-sanctioned access: Because some jurisdictions allow their authorities, under certain circumstances, to access all data stored on servers located in that jurisdiction (e.g., in the United States under provisions of the *Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act,* known as USA PATRIOT Act: see the Treasury Board Secretariat 's overview), the researcher must ensure that all the databases containing any information related to the survey are stored on **servers and back- up servers** located solely in Canada.

   a) If the client has first consented in writing, these servers may be located in another country where:

      i. equivalent protections are given to personal information as in Canada under legislation such as the *Privacy Act*, R.S. 1985, c. P-21, and the *Personal Information Protection and Electronic Documents Act*, S.C. 2000, c. 5, and under any applicable policies of the Government of Canada; and

ii. the laws do not allow the government of that country or any other entity or person to seek or obtain the right to view or copy any information relating to the survey without first obtaining the client's written consent.

b) In connection with giving consent to locating a database in another country, the client may, at their option, require the researcher to provide a legal opinion (from a lawyer qualified in the foreign country) that the laws in that country meet the requirements of paragraph 2) a), above, or may require the researcher to pay for the Government of Canada to obtain such a legal opinion. The Government of Canada has the right to reject any request to store survey data in a country other than Canada if there is any reason to be concerned about the security, privacy, or integrity of the data. The Government of Canada may also require that any data sent or processed outside of Canada be encrypted with Government of Canada-approved cryptography and that the private key required to decrypt the data be kept in Canada in accordance with key management and storage processes approved by the Government of Canada.

c) The researcher must ensure that all servers, including back-up servers, on which any data relating to the survey is stored are physically and logically independent (meaning there is no direct or indirect connection of any kind) from all other databases, unless those databases are located in Canada (or in another country approved by the client under paragraph 2) a)) and otherwise meet the requirements of this section.

d) The researcher must ensure that all data relating to the survey is processed only in Canada or in another country approved by the client under paragraph 2) a).

e) The researcher must ensure that all domestic network traffic (meaning traffic or transmissions initiated in one part of Canada to a destination or individual located in another part of Canada) is routed exclusively through Canada, unless the client has first consented in writing to an alternate route. The client will only consider requests to route domestic traffic through another country that meets the requirements of paragraph 2) a).

Questions Addressed by the Panel

The Panel considered whether any other standards are required with respect to cloud-based storage, either in terms of location of servers/back-up servers or any other aspects of data security specific to cloud-based storage.

Advisory Panel Rsponse

For the most part there were no suggested changes to the current standards. This is a complex area: it requires expertise in the legal and regulatory framework affecting data access and use not only in Canada but in other countries as well where servers might be located, and it requires an understanding of GC policies in this area. The Panel did not consider itself to be experts in these areas.

There was some discussion about allowing cloud storage in countries with equivalent standards. The current standards permit this providing certain conditions are met as detailed in 2a and 2b. However, it was suggested that applying 2b on a case by case basis is cumbersome and time consuming. A better approach would be for the GC to have a pre-approved list of countries that satisfy the conditions set out in 2a and are acceptable for cloud storage of GC POR data. It was

noted that this is an approach taken by the European Commission with respect to protection of personal information in non-EU countries[28].

Some other points to note:

- It is unclear what the privacy and security status is of servers in Canada owned by non-Canadian companies. Does the other country have legal rights to access the data stored on the servers in Canada? If so, does that mean that not only must the server be located in Canada, but it must also be owned by a Canadian company? That would require amending the current standards.

  This question is raised by the recent passage in the U.S. of the CLOUD Act, which allows federal U.S. law enforcement to compel U.S.-based technology companies to provide requested data stored on servers regardless of whether the data are stored in the U.S. or in another country.

- The impact of the current standards goes well beyond simple storage of data. Cloud-based or cloud-integrated apps as a way of providing and accessing software is a rapidly growing business model. Data storage and/or use is one component of these apps. As a result, the current standard blocks usage of cloud-based apps providing research services that collect or access respondent data if these apps use non-Canadian servers (and beyond that, potentially if they are owned by a non-Canadian company – as referenced above).

[28] *Adequacy of the protection of personal data in non-EU countries*. https://ec.europa.eu/info/law/law-topic/data-protection/data-transfers-outside-eu/adequacy-protection-personal-data-non-eu-countries_en

Surveys and Social Media

Background and Questions

A document relevant to this topic is MRIA's *Appendix C – Guideline on Social Media Research*. Note however that the Panel focus was limited to usage of social media only in connection with conducting online or telephone surveys, and that also meet PORD's definition of public opinion research[29]. Under this definition (1) there must be attitudinal/opinion questions in the research, and (2) the research must be based on asking questions. The MRIA document deals with social media research more generally, not just survey research.

The *Introduction* to the MRIA's *Appendix C – Guideline on Social Media Research* – states:

> The concept of consumers generating their own content on the internet has become ubiquitous. This has created new opportunities for researchers to observe, interact and gather information. Many techniques have been developed to leverage social media such as community panels, crowd-sourcing, co-creation, netnography, blog mining and web scraping. It is likely that many more will evolve over the coming years as the Internet continues to change.

Many of the research possibilities referred to above fall outside the scope of PORD's Online and Telephone Standards, because the activities do not qualify as public opinion research surveys. However, particularly in the case of market research online communities (MROCs), it is possible to do an online survey and perhaps even a telephone survey. And, it is possible that other types of social media venues can be platforms for sampling and for administering surveys.

If an online or telephone survey is done using a social media venue as the sample source, and perhaps additionally as the medium for administering an online survey, the research project would have to conform to all of the relevant standards, that is, the Online Survey Standards or the Telephone Survey Standards.

The question is whether anything needs to be added to the standards to cover online or telephone surveys that make use of social media – meaning, the sample is sourced from a social media, and the additional possibility that the survey is administered via the social media venue.

One perspective is that nothing really needs to be added to the standards for social media-based surveys. The Standards lay out requirements for proposal documentation, questionnaire design, sampling, retaining public confidence, data collection, data security, etc. Adhering to these standards could be considered sufficient for an acceptable social media-based survey conducted for the Government of Canada.

A different perspective is that there may be some issues specific to social media-based surveys that are not clearly covered by the Standards.

[29] PORD's definition of public opinion research can be found at: https://www.canada.ca/en/treasury-board-secretariat/services/government-communications/public-opinion-research-government.html

The MRIA's *Appendix C – Guideline on Social Media Research* covers social media research broadly, not just surveys. The document describes "key principles" for researchers. These principles are consistent with both the Online and Telephone Standards, but are explained specifically in terms of application to social media:

2.1 Distinguishing market research as the purpose
2.2 Conforming to the law
2.3 Consent and notification
2.4 Protecting identifiable data
2.5 Ensuring no harm
2.6 Children, young people and persons with functional cognitive disability
2.7 Reputation of the industry
2.8 Reporting

Also of particular interest for the survey standards is Section 3 of the MRIA document, *Recommendations for Specific Social Media*, and within this, Section 3.2 *Private social media area issues* and Section 3.3 *Market research social media area issues.* These address aspects of permission, informed consent and privacy specific to these types of social media venues used for research.

Questions Addressed by the Panel

The Panel considered whether there are any additional standards required for surveys that use a social media venue as a sample source or to administer a survey.

Advisory Panel Response

No additional standards are needed for surveys that use a social media venue as a sample source or to administer a survey. The current standards, together with the various changes recommended elsewhere by the Panel, are sufficient to ensure that any such surveys meet the quality requirements for GC POR surveys.

MRIA's *Appendix C – Guideline on Social Media Research* on social media states the following principles – all of which are already addressed in the current standards.

- *2.1 Distinguishing market research as a purpose:*
  - Section 2.1 of the GC Standards, *Questionnaire Design*, requires description of the purpose and sponsor of the survey and of the research firm
  - Section 5.1 of the GC Standards, *Respondent Rights*, requires getting informed consent
- *2.2 Conforming to the law:* This is already baked into the GC Standards, and as well can be taken as a given.
- *2.3 Consent and notification:* As noted, Section 5.1 of the GC Standards requires informed consent. Section 5.1.5, *Use of Respondent Lists*, requires lists to be permission-based for research purposes.
- *2.4 Protecting identifiable data:* This is addressed in Section 5.1.4 of the GC Standards, *Protection of Anonymity and Confidentiality.* Also relevant is Section 5.3, *Privacy Issues*

Specific to Online Survey Research, which requires disclosure of privacy-related matters specific to online research.

Note that respondent-supplied images (photos, videos) and recordings are not specifically addressed in the current standards, but see Question A3.1 in this module for a possible standard to explicitly address these types of potentially personally identifiable information.

- *2.5 Ensuring no harm:* This section of MRIA's Appendix C seems to overlap quite a bit with their 2.4 above. In any event, also relevant here in the current GC Standards are Section 5.1.1, *Informed Consent and Use of Information*, which forbids non-research uses of survey data (unless explicit consent is provided), and Section 5.2, *Avoidance of Harassment*.
- 2.6 *Children, Young People and Persons with Functional Cognitive Disability:* This is addressed in Section 6 of the GC Standards, *Data Collection from Children, Young People or Vulnerable Respondents*, and indeed seems to us be more thoroughly addressed here than in the MRIA document.
- *2.7 Reputation of the Industry:* GC Standards are high, and nothing further needs to be said to protect the "reputation of the industry."
- *2.8 Reporting:* Extensively addressed in the GC Standards – in particular Section 4, *Sampling Procedures*, and Section 14/15, *Mandatory Survey Report Requirements*.
- In MRIA's Appendix C, *3.2 Private social media issues,* and *3.3 Market research social media area issues* deal more with the management of these types of sites as well as non-survey information gathering, and these matters are not the subject domain of GC Online & Telephone Survey Standards.

If anything were to be added to the survey standards pertaining to social media, it would be more in the way of reminders to take care to follow the standards on matters where there is concern that use of social media may be problematic. For example, depending on the social media venue being used, there may be particular concern that one is dealing with "real people" (i.e. not bots) or that the people are who they say they are. In this case, one could add to Section 4, *Sampling Procedures*, something like the following: *if the research involves sourcing sample from a social media platform, the researcher should outline the additional measures they have taken to confirm respondent eligibility and authenticity*.

Accessibility and Literacy

Background and Questions

The Online and Telephone Standards do not contain any standards or guidelines pertaining to accessibility.

The Web Accessibility Initiative (WAI) home page (https://www.w3.org/WAI/intro/usable) provides definitions of accessibility, usability and inclusion:

> **Accessibility:** *Accessibility* addresses discriminatory aspects related to equivalent user experience for people with disabilities, including people with age-related impairments. For the web, accessibility means that people with disabilities can perceive, understand, navigate, and interact with websites and tools, and that they can contribute equally without barriers.
>
> **Usability:** *Usability* and *user experience design* is about designing products to be effective, efficient, and satisfying. Specifically, ISO defines usability as the "extent to which a product can be used by specified users to achieve specified goals effectively, efficiently and with satisfaction in a specified context of use."
>
> **Inclusion:** *Inclusive design*, *universal design*, and *design for all* involves designing products, such as websites, to be usable by everyone to the greatest extent possible, without the need for adaptation. Inclusion addresses a broad range of issues including access to and quality of hardware, software, and Internet connectivity; computer literacy and skills; economic situation; education; geographic location; and language — as well as age and disability.

Accessibility for Online Surveys

According to PORD, the Treasury Board Secretariat (TBS) is working on a proposed policy for accessibility standards specific to all devices used to access online surveys. The results of this development work will probably be available in a year or so. When the TBS policy is finalized, it will take precedence.

Note that the Panel made recommendations for standards and guidelines for mobile-friendly questionnaires, and these are to some extent relevant to accessibility of online surveys.

On pages 25-28 of the literature review commissioned by PORD there is a section on *Accessibility*. In this section, there are three lists of guidelines from various sources. In the literature review, they are associated with *making mobile surveys* accessible, but many of the items pertain to online surveys generally. The three lists are as follows:

> **Page 26: Web Content Accessibility Guidelines:**
>
> - providing text alternatives for any non-text content which allows it to be changed into other forms, such as large print, braille, or speech;
> - creating content that can be presented in different ways without losing information or structure;
> - separating the foreground from the background so it is easier to see and hear content;
> - making all functionality available through a keyboard;

- making pages appear and operate in predictable ways;
- maximizing compatibility with user agents, including assistive technologies, such as screen magnifiers, visual reading assistants, screen readers with synthesized speech or braille, text-to-speech software, speech recognition software, alternative keyboards and alternative pointing devices

Pages 26-27: W3C's Web Accessibility Initiative "that could apply to survey research":

- Avoiding features that require the user to input text rather than selecting from values. In the context of survey research, this means using open-ended questions sparingly.
- Minimizing the amount of information on each page because the screens of mobile devices are typically smaller than those of desktop and laptop computers. For survey research, this means one question per screen. Large grid style questions, therefore, are best avoided.
- Ensuring form fields are presented below rather than beside their labels because screens are smaller, and scrolling side-to-side is not as convenient as scrolling down. In the context of survey research and design, this means positioning a field to enter data, such as a text box for an open-ended question or a box to enter date of birth, underneath the question.
- Ensuring text is resizable up to 200 percent, without the use of assistive technology.
- Avoiding high-resolution images that use up bandwidth (which can be expensive depending on a user's data plan and/or time consuming to access in an area where high speed internet access is not accessible). This could apply to the images used to brand a mobile survey or that are presented to respondents for reaction (e.g., ad testing).

Page 27 "commercial tools" list:

- JavaScript-based questions are best avoided because they tend to not work well with screen readers. This means some of the more interactive and visually engaging types of questions, depending on the survey tool, should not be used for surveys that must be accessible. This includes, for example, drag and drop questions (often used for ranking items), slider questions (often used for rating scale questions), heat map questions, and questions designed to sum the numeric data entered by respondents.
- Make questions as concise as possible. For respondents using screen readers, an excessively wordy question will be tedious to listen to (just as it would be when read by a telephone interviewer). If the question is too long, or there are too many response options, the respondent may have difficulty answering the question and/or provide an unintended response.
- Use grid style questions sparingly. In fact, rather than use a table/matrix to concisely present items/statements for a respondent to rate, it is best to opt for a survey with more screens/pages, and give each item/statement to be rated its own screen/page.
- Make sure warning messages in a survey can be read by screen readers.
- Make the layout as simple as possible and use one screen/page per question, which is more user-friendly for anyone using a screen reader.

Accessibility for Telephone Surveys

There are currently no accessibility standards in the Telephone Standards. There are ways to make telephone surveys more accessible, such as providing options for TTY or alternative modes such as online or mail, and allowing use of proxy respondents.

Questions Addressed by the Panel

The Panel considered whether a statement should be added to the Standards about the importance of the principles of accessibility – including literacy considerations, usability and inclusion, and what if any specific guidelines might be provided for online and telephone surveys.

Advisory Panel Response

The majority of Panelists supported stating a general guideline encouraging accessibility, followed by some examples of steps that could be taken to improve accessibility in online or telephone surveys. These examples would be guidelines, not requirements. The general guideline could be:

> All Government of Canada online/telephone surveys should strive to be as accessible as possible for all respondents. Research suppliers are expected to work closely with the client team to accommodate the needs of a respondent for whom standard data collection practices do not allow for an equivalent or adequate participation in a research study for which they would otherwise qualify.

Examples could be helpful to clients and suppliers, and could include:

> **[After the general statement:]** There are many aspects to survey accessibility. Some examples to consider include:
>
> **Online surveys:**
> - Maximize compatibility with user agents, including assistive technologies, such as screen magnifiers, visual reading assistants, screen readers with synthesized speech or braille, text-to-speech software, speech recognition software, alternative keyboards and alternative pointing devices
> - Ensure text is resizable without the use of assistive technology
> - Make all functionality available through a keyboard
> - Make sure warning messages can be read by screen readers
> - Make the layout as simple as possible and one screen per question, which is more user-friendly for anyone using a screen reader
>
> **Telephone surveys:**
> - Provide alternative survey modes for people who cannot use a telephone
> - Allow a person to designate a proxy to act as the respondent

Several Panelists had somewhat different recommendations:

- A few Panelists suggested stating only the general guideline encouraging accessibility, but not giving examples. The reasons for not giving examples were (a) the list is not exhaustive, (b) the list may quickly become dated as technology evolves, and (c) the list would be superseded by the upcoming Treasury Board guidance.
- One Panelist favoured the option preferred by the majority, but with one change: the general statement would be worded as a requirement rather than a guideline in order to emphasize the importance of accessibility. A counterargument to this was that because of the numerous methods by which accessibility might be addressed in any given survey, and because of the varying relevance and feasibility of these methods for any given survey, it would be unclear how to implement and enforce a general requirement.

Appendix A: Background and Questions – Discussion Board #1

DISCUSSION BOARD #1: BACKGROUND
(Start date – April 16; End date – May 4)

Topics:

The following contains background information and the questions for the group discussion. More questions may be developed for this Discussion Board as it progresses and based on the views and suggestions of Panel members.

Please note the following:

1) **You will need to refer to this Background for the detailed commentary on the questions shown on the Discussion Board.** The Discussion Board displays only the questions. This Background also provides, for example:
   -- specific excerpts from relevant sources
   -- references to information available

   This type of detail has **not** been included in the Discussion Board questions.
2) Please answer **all** the relevant questions; it is important to know the point of view of each member of the Advisory Panel.
3) Wherever possible, we need to have as detailed answers as possible, particularly in areas where there is possible disagreement between Panel members on either a principle stated, or the language used to express the principle.
4) We need all members of the panel to begin posting answers to the questions as soon as possible and shortly after the Discussion Board is up and running. Otherwise, there will be

little if any time left to debate issues, to try to get clarification or more detail from the Panel, or to come to some agreement on issues where Panel members have different views.

It is the mandate of the Advisory Panel ***to reach consensus where possible even though it is not an essential outcome of deliberations***. Where there is initial disagreement, please talk to one another to see if a consensus can be found. Think of lack of consensus as a last resort.

5) Just to remind you about the terminology being used:

Standards	Practices which are ***requirements*** for all quantitative research conducted by the Government of Canada
Guidelines	Practices which are ***recommended,*** but would not be requirements; that is, known good practices or criteria that serve as a checklist to ensure quality research but are not necessarily applied to every study

Wherever possible for the sake of simplicity, the standards or guidelines should be the same for Government of Canada POR telephone and online surveys.

6) PORD asks that the following criteria be applied when formulating standards and guidelines:

In general, we wish to adhere to the following criteria when revising/developing Standards:

- *Be measurable*
- *Not conflict with other GC departments (e.g., TBS) or related policies or standards (e.g., IT, legal)*
- *Avoid material that will quickly become out-of-date*
- *Not overly prescriptive*
- *Minimize web links in the content*

7) You will see various citations in the footnotes of this Background. We have tried to make the Background "self-sufficient" so that it is not necessary to read the cited articles. However, if you would like to get any of these documents, let us know which ones you would like and we'll email them to you.

8) The following are other documents Panel members will need to refer to, and so are being provided to members:

- The Online and Telephone Standards
- For Section C of the discussion board: *AAPOR Guidance on Reporting Precision for Nonprobability Samples*
- For Section D, an article, *It's Time for Pollsters to Report Margins of Error More Honestly*
- For Section F, an excerpt from the report of the 2008 *Advisory Panel on Online Public Opinion Survey Quality*

We're also sending the literature review commissioned by PORD.

A. Definitions of Probability versus Non-Probability Samples

Section 4 of the *Standards* is *Sampling Procedures,* and it includes standards for both probability sampling and non-probability sampling. Section 4 does not give definitions of these two types of sampling procedures. This section may be revised to include definitions and some examples to help users determine whether a sample is a probability sample or a non-probability sample. The objectives is to formulate appropriate definitions and examples to include in the Standards.

Probability sampling: Statistics Canada[30] states:

> Probability sampling is a method of sampling that allows inferences to be made about the population based on observations from a sample. In order to be able to make inferences, the sample should not be subject to selection bias. Probability sampling avoids this bias by randomly selecting units from the population (using a computer or table of random numbers). It is important to note that random does not mean arbitrary. In particular, the interviewers do not arbitrarily choose respondents since then sampling would be subject to their personal biases. Random means that selection is unbiased – it is based on chance. With probability sampling, it is never left up to the discretion of the interviewer to subjectively decide who should be sampled.
>
> There are two main criteria for probability sampling: one is that the units be randomly selected, the second is that all units in the survey population have a non-zero inclusion probability in the sample and that these probabilities can be calculated. It is not necessary for all units to have the same inclusion probability, indeed, in most complex surveys, the inclusion probability varies from unit to unit. (p.91)

Non-probability sampling: Some excerpts from the same source as above:

> Non-probability sampling is a method of selecting units from a population using a subjective (i.e., nonrandom) method. Since non-probability sampling does not require a complete survey frame, it is a fast, easy and inexpensive way of obtaining data. The problem with non-probability sampling is that it is unclear whether or not it is possible to generalize the results from the sample to the population. The reason for this is that the selection of units from the population for a non-probability sample can result in large biases.
>
> Due to selection bias and (usually) the absence of a frame, an individual's inclusion probability cannot be calculated for non-probability samples, so there is no way of producing reliable estimates or estimates of their sampling error. In order to make inferences about the population, it is necessary to assume that the sample is representative of the population. This usually requires assuming that the characteristics of the population follow some model or are evenly or randomly distributed over the population. This is often dangerous due to the difficulty of assessing whether or not these assumptions hold.

[30] Statistics Canada, 2003, *Survey Methods and Practices*, Catalogue no. 12-587-X

Proposed revision to Section 4.1 of the Telephone & Online Standards

The following is the proposed revision. The brown text is the proposed new text.

4. SAMPLING PROCEDURES

4.1. General

All researchers must:

1) Clearly state the target group (universe) definition for the research study; in the case of Internet surveys this includes explicit identification of whether or not non-Internet users are part of the target group definition;
2) Clearly state the method(s) used to obtain a sample of this target group, including whether the method is a probability survey, a non-probability survey, or a census. Definitions and examples of each method are as follows:
   a) Probability sample: respondents are randomly selected from the survey's target population, and each respondent's probability of inclusion can be calculated. Probability sampling is a method for obtaining a sample projectable to the target population.

      Some examples:
      - Random-digit-dial (RDD) telephone survey of Canadians.
      - Random sampling from a list of all members of the target population.
      - Random sampling from a panel that is itself a probability sample of the target population.
      - Website intercept survey in which target population is visitors to the website, and visitors are randomly sampled to take part in a survey.

   b) Non-probability sample: a sample that does not meet the requirements of a probability sample – that is, respondents are not randomly selected from the survey's target population, and/or each respondent's probability of inclusion cannot be calculated. Additional steps must be taken to try to make results from a non-probability sample representative of the target population.

      Some examples:
      - A sample drawn from a research panel consisting of people who volunteer to join the panel and do surveys. Note that a sample collected using probabilistic methods from sampling frames that were compiled using non-probability methods is considered a non-probability sample.
      - Quota sampling, in which the selection of respondents is based on judgment, convenience or some other nonrandom process.

   c) Census: An attempt is made to collect data from every member of the target population. Note that a census can be subject to other types of survey error, notably coverage error and nonresponse, so not every member of the target population may be in the final data set.

Questions Posted on the Discussion Board

A1. Do you have any suggestions for changes to the proposed revision to Section 4.1 #2?

B. Maximizing Representativeness of Non-probability Surveys

Surveys based on non-probability sampling have become more common in marketing research, particularly because of the growth of online opt-in panels that provide significant cost savings over telephone probability samples. Public opinion research surveys for the Government of Canada have historically usually used probability sampling, but there may be more usage of non-probability surveys if there is confidence that these can deliver results that are representative of the target population being surveyed.

The objective is to revise the Standards to emphasize the importance of striving for representativeness in non-probability surveys, and to explain in the proposal how this will be done.

The current standards address this objective to some extent, but the intent is to make the requirement more explicit and detailed.

Current Standards

Standard #1 *Proposal Documentation*, does not contain any explicit language on the importance of maximizing representativeness. Standard 1.2.2 *Sample/Sampling Details* only says:

- State if census, probability or non-probability
- If non-probability, then *provide rationale for choosing a non-probability sample*
- 1.2.2.5 makes a reference to "quota controls" – *Define respondent eligibility/screening criteria, including any quota controls*

There is some relevant language in Standards 4.3 *Non-Probability Sampling* and 4.4 *Quota Sampling*

Standard 4.3.2 *Sampling for Non-probability Samples*

1) As for probability sampling, the list or sample source must be stated, including its limitations in covering the universe for the target sample.
2) The precise quota control targets and screening criteria must also be stated including the source of such targets (e.g., census data or other data source).

Standard 4.4 *Quota Sampling*

Quota sampling techniques are typically used for panel surveys and personal intercept studies to achieve sample representativeness. Quotas may also be used to control representativeness on other data collection methodologies.

1) A full description of the regional, demographic or other classification variable controls used for balancing the sample to achieve representativeness must be described.
2) The precise quota control targets and screening criteria must also be stated including the source of such targets (e.g., census data or other data source).

Pew Research has recently published two studies[31] that examined factors affecting the extent to which results from a non-probability survey can be representative[32]. Some observations on their results:

- There are steps one can take that will improve the representativeness of non-probability survey results. It is also the case that some degree of bias will likely remain. There is no "magic method" that has been shown to eliminate all bias.
- The magnitude of bias can vary across different measures in the survey, and can vary substantially. This implies it is important to identify the key measures given the survey's purpose when making decisions about steps to take to make the survey results more representative. This is because a factor used to try to enhance representativeness will only be effective if it is correlated with the survey variables of interest.
- When selecting factors to enhance the representativeness of the sample or to make post hoc adjustments (e.g. weighting), it is better to avoid using only a small number of factors (e.g. avoid simple schemes such as gender, age and region).
- Incorporating nondemographic factors such as attitudes or behaviour correlated with the key survey measures of interest can improve representativeness providing there are high quality population statistics for such factors to use as benchmarks.
- The factors/variables selected to make adjustments tend to be more important for improving representativeness than "complex statistical methods" such as propensity weighting or sample matching. While the choice of statistical method can certainly matter, appropriate choice of adjustment variables is a critically important first step.

The following is an excerpt from summary section of the 2018 Pew Research report, *For Weighting Online Opt-In Samples, What Matters Most?*

> ***For Weighting Online Opt-In Samples, What Matters most?*** *The right variables make a big difference for accuracy. Complex statistical methods, not so much.*
>
> A growing share of polling is conducted with online opt-in samples. This trend has raised some concern within the industry because, while low participation rates pose a challenge for all surveys, the online opt-in variety face additional hurdles. By definition they do not cover the more than 10% of Americans who don't use the internet. The fact that potential respondents are self-selected means that there is still substantial risk that these samples will not resemble the larger population. To compensate for these challenges, researchers have employed a variety of statistical techniques, such as raking, propensity weighting and matching, to adjust samples so that they more closely match the population on a chosen set of dimensions. Researchers working with online opt-in samples must make a great many decisions when it comes to weighting. What factors should guide these decisions, and which ones are most consequential for data quality?
>
> A new Pew Research Center study adds to the survey field's broader efforts to shed light on these questions. The study was based on over 30,000 online opt-in panel interviews conducted in June and July of 2016, with three vendors, and focuses on national (as opposed to state or local level) estimates. We evaluated three different weighting techniques, raking, propensity weighting and matching, both on their own and in combination. Every method was applied using two sets of adjustment variables: basic

[31] Pew Research Center, May 2016, "Evaluating Online Nonprobability Surveys."
Pew Research Center, January 2018, "For Weighting Online Opt-In Samples, What Matters Most?"
[32] Another interesting study that draws similar conclusions is the following, which compared non-probability surveys to "low-response rate" probability telephone surveys: Dutwin, D. & Buskirk, T., 2017, "Apples to Oranges or Gala versus Golden Delicious? Comparing data quality of nonprobability internet samples to low response rate probability samples", *Public Opinion Quarterly,* Vol. 81, Special Issues, 2017, pp. 213-249

demographics (age, sex, race and ethnicity, education, and geographic region), and a more extensive set that included both demographics and a set of variables associated with political attitudes and engagement (voter registration, political party affiliation, ideology and identification as an evangelical Christian). Each procedure was performed on simulated samples ranging in size from n=2,000 to n=8,000.

The procedures were primarily appraised according to how well they reduced bias on estimates from 24 benchmark questions drawn from high-quality federal surveys. They were also compared in terms of the variability of weighted estimates, accuracy among demographic subgroups, and their effect on a number of attitudinal measures of public opinion.
Among the key findings:

- **Even the most effective adjustment procedures were unable to remove most of the bias.** The study tested a variety of elaborate weighting adjustments to online opt-in surveys with sample sizes as large as 8,000 interviews. Across all of these scenarios, none of the evaluated procedures reduced the average estimated bias across 24 benchmarks below 6 percentage points – down from 8.4 points unweighted. This means that even the most effective adjustment strategy was only able to remove about 30% of the original bias.
- **When it comes to accuracy, choosing the right variables for weighting is more important than choosing the right statistical method.** Adding a set of politically focused variables to the weighting adjustment reduced the average estimated bias by an additional 1.4 percentage points relative to adjusting only on basic demographics (e.g., age, education, race). While that might seem small, a difference of 1.4 points in the average implies that about 36 percentage points of bias were removed overall, but spread out across all 24 variables. Benchmarks most strongly associated with the political adjustment variables saw the largest improvements. In contrast, the use of more complex statistical methods never reduced the average estimated bias by than 0.3 points beyond what was achieved with raking, the most basic statistical method evaluated.
- **The benefits of adding political variables to adjustment differ by survey topic[33].** Perhaps not surprisingly, benchmarks related to political engagement saw the largest improvement with the addition of political adjustment variables. Unweighted, these benchmarks had an average estimated bias of 22.3 percentage points, more than any other topic. While demographic weighting reduced the average bias by an average of 2.9 points, the effect of adding political adjustment variables was four times as large, reducing bias by 11.7 points and cutting the average estimated bias nearly in half (to 10.6 percentage points). Benchmarks pertaining to civic engagement and technology use also benefited disproportionately from political adjustment variables, though to a lesser degree. For benchmarks related to family composition and other personal characteristics, variable selection made little difference and proved mildly detrimental for questions of personal finance.
- **The most basic weighting method (raking) performs nearly as well as more elaborate techniques based on matching.** When weighting on both demographic and political variables, methods based on matching resulted in the lowest average bias across the full set of 24 benchmarks – either in combination with raking at smaller sample sizes (n=less than 4,000) or on its own when the sample size was larger. Even so, procedures that only used raking (the least complex method evaluated) performed nearly as well, coming in 0.1 to 0.3 points behind the most effective method, depending on sample size. For benchmarks related to political engagement, the benefits from the more complex approach are somewhat larger than for other topics, doing between 0.5 and 1.2 points better than raking depending on sample size, but nowhere near the magnitude of improvement derived from weighting on political variables in addition to demographics. If the data necessary to perform matching are readily available and the process can be made routine, then a combination of matching and other methods like raking is likely worthwhile, providing incremental but real improvements. In other situations, such marginal improvements may not be worth the additional statistical labor.
- **Very large sample sizes do not fix the shortcomings of online opt-in samples.** While an online opt-in survey with 8,000 interviews may sound more impressive than one with 2,000, this study finds virtually no difference in accuracy. When adjusting on both demographic and political variables, the most effective procedure at n=8,000 was only 0.2 points better than the most effective procedure at

[33] Because of the nature of many of Pew Research's interests, "political variables" are relevant to their work. The key take-away here is that appropriately chosen attitudinal or behavioural measures can help reduce bias. The nature of these variables, including whether or not they are "political", will depend on the survey topic.

n=2,000. While a large sample size may reduce the variability of estimates (i.e., the modeled margin of error), this is of little help from a "total survey error" perspective. For example, raking on demographic and political variables, the average modeled margin of error across all 24 benchmark variables is ±1.8 percentage points when n=2,000 and ±0.5 points when n=8,000, but the average bias holds steady at 6.3 points. As the sample size increases, estimates become less dispersed and more tightly clustered, but they are often more tightly clustered around the wrong (biased) value.

The weighting procedures tested in this report represent only a small fraction of the many possible approaches to weighting opt-in survey data. There are a host of different ways to implement matching and propensity weighting, as well as a variety of similar alternatives to raking (collectively known as calibration methods). We also did not evaluate methods such as multilevel regression and poststratification, which require a separate statistical model for every outcome variable. Add to this the innumerable combinations of variables that could be used in place of those examined here, and it is clear that there is no shortage of alternative protocols that might have produced different results.

But whatever method one might use, successfully correcting bias in opt-in samples requires having the right adjustment variables. What's more, for at least many of the topics examined here, the "right" adjustment variables include more than the standard set of core demographics. While there can be real, if incremental, benefits from using more sophisticated methods in producing survey estimates, the fact that there was virtually no differentiation between the methods when only demographics were used implies that the use of such methods should not be taken as an indicator of survey accuracy in and of itself. A careful consideration of the factors that differentiate the sample from the population and their association with the survey topic is far more important.

Proposed revisions to the Telephone & Online Standards

Revisions are proposed for Section 1, *Proposal Documentation*, and Section 4, *Sampling*. The brown text is new/revised material.

Proposed revision to Section 1.2.2: *Proposal Documentation, Sample/Sampling Details*

1.2.2. Sample/Sampling Details

1) Provide details related to target population:
   a) the definition of the target population in terms of its specific characteristics and geographic scope, including the assumed incidence of the population and any key sub-groups and how the incidence was determined/obtained (e.g., supplied by the client);
   b) the total sample size and the sample sizes of any key sub-groups.
2) Describe the sampling procedures, including:
   a) the sample source;
   b) the sample frame;
   c) whether a sample survey or a census will be conducted and, if a sample, whether probability or non-probability sampling will be applied (see section 4 for additional information to include in the proposal).
3) Explain respondent selection procedures.
4) Indicate the number of re-contact attempts and explain re-contact attempt procedures.
5) Define respondent eligibility/screening criteria, including any quota controls.
6) For non-probability samples, provide the rationale for choosing a non-probability sample. If the survey results will be used to make statements about a population, steps must be taken to maximize the representativeness of the sample with respect to the target population, and these steps must be documented in the research proposal and in the survey report (see section 4.3).

Proposed revision to Section 4.3, *Non-Probability Sampling*

Note: This proposal includes eliminating Section 4.4, *Quota Sampling*, and moving relevant content to Section 4.3. *Non-Probability Sampling:*

For reference, Section 4.4, *Quota Sampling*, is:

Quota sampling techniques are typically used for panel surveys and personal intercept studies to achieve sample representativeness. Quotas may also be used to control representativeness on other data collection methodologies.

1) A full description of the regional, demographic or other classification variable controls used for balancing the sample to achieve representativeness must be described.
2) The precise quota control targets and screening criteria must also be stated including the source of such targets (e.g., census data or other data source).
3) Deviations from target achievement must be shown in the report (i.e., actual versus target).

4.3. Non-Probability Sampling

4.3.1. Justification of Use of Non-probability Samples

1) When a choice is made to use a non-probability sample, that choice must be justified, in both the research proposal and the research report. The justification must take into account the statistical limitations in reporting on data from a non-probability sample, and limitations in generalizing the results to the target population.

4.3.2. Sampling for Non-probability Samples

1) As for probability sampling, the list or sample source must be stated, including its limitations in covering the universe for the target sample.
2) If the survey results will be used to make statements about a population, steps must be taken to maximize the representativeness of the survey results with respect to the target population, and these steps must be documented in the research proposal and in the survey report.

   These steps include:
   - Controls on sample composition to maximize representativeness, such as quota sampling
   - Weighting
3) A full description must be provided of the regional, demographic or other classification variables used to maximize the representativeness of the sample and survey results. In selecting variables, also consider their likely correlation with key survey measures (adjustment variables that are uncorrelated with survey measures will do little to improve representativeness). Behavioural or attitudinal variables can also improve representativeness, providing relevant, high quality benchmarks exist for the target population.
4) The precise quota control targets and screening criteria must also be stated including the source of such targets (e.g., census data or other high quality data source).
5) Deviations from target achievement must be shown in the report (i.e., actual versus target).

Questions Posted on the Discussion Board

B1. Do you have any suggestions for changes to the proposed revisions to Section 1.2.2 (*Proposal Documentation; Sample/Sampling Details*) or Section 4.3.2 (*Sampling; Sampling for Non-probability Samples*)?

B2. Are there any revisions to Section 1 *Proposal Documentation* or to Section 4 *Sampling* to address the objective to revise the Standards to emphasize the importance of striving for representativeness in non-probability surveys, and to explain in the proposal how this will be done?

C. Statistical Treatment of Non-probability Survey Results

The issue here is what can and should be done with respect to the statistical treatment of non-probability surveys.

The current standards are in section 4.3.3, *Statistical Treatment of Non-probability Samples*. Note that this section comes after standards pertaining to justification of the use of a non-probability sample (4.3.1), and requirements for information disclosure and maximizing representativeness (section 4.3.2, which was the subject of the previous question to the Panel).

Current Standards

4.3.3. Statistical Treatment of Non-probability Samples

1) There can be no statements made about margins of sampling error on population estimates when non-probability samples are used.

2) The survey report must contain a statement on why no margin of sampling error is reported, based on the following template: "Respondents for this survey were selected from among those who have [volunteered to participate/registered to participate] in online surveys. The results of such surveys cannot be described as statistically projectable to the target population. [If weighting was done, state the following sentence on weighting:] The data have been weighted to reflect the demographic composition of (target population). Because the sample is based on those who initially self-selected for participation [in the panel], no estimates of sampling error can be calculated."

   This statement must be prominently placed in descriptions of the methodology in the survey report, including the executive summary.

3) In general, for non-probability surveys it is not appropriate to use statistical significance tests. However, tests of significance with non-probability samples are appropriate when the objective is to establish the extent of the relationship among variables. If tests of significance are used with non-probability samples, it must be clearly noted that conclusions from these tests cannot be generalized to any population.

   Any use of descriptive statistics must clearly indicate that they are not formally generalizable to any group other than the sample studied, and there cannot be any formal statistical inferences about how the descriptive statistics for the sample represent any larger population.

   In the case of non-probability surveys that employ an experimental design in which respondents are randomly assigned to different cells, it is appropriate to use statistical significance tests to compare results from different cells.

The 4.3.3 standards are consistent with, albeit more detailed than, the MRIA standards (note that these apply to research results generally, not just non-probability samples)[34]:

MRIA Code of Conduct

"Researchers must not present research results with greater confidence than the data warrants. Instead, as responsible professionals, members must point out the relevant limitations of the research. This includes but is not limited to the following guidelines:

[34] MRIA Code of Conduct for Market and Social Research, Appendix L – Polling Standards for the Canadian Marketplace, June 2017

i. Disclosing relevant potential sources of error, both sampling and non-sampling (e.g. response, non-response, measurement, coverage, etc.).

ii. Being explicit about the assumptions made about data accuracy when employing quota or stratification methods with probability samples.

iii. Refraining from making unqualified statements about confidence intervals or margins of sampling error on population estimates when probability samples are not used. For example, panels of repeat volunteers will not ordinarily qualify as sources of probability samples of the general population."

Until 2015, the American Association for Public Opinion Research (AAPOR) took a similar stance. However, in 2015 AAPOR revised its *Code of Professional Ethics and Practices* to allow for reporting of measures of precision from non-probability samples.

The change was apparently motivated in part by the 2013 *Report of the AAPOR Task Force on Non-probability Sampling*, which stated:

> We believe that users of non-probability samples should be encouraged to report measures of the precision of their estimates, but suggest that, to avoid confusion, the set of terms be distinct from those currently used in probability sample surveys. The precision of estimates from non-probability samples is not the average deviation over all possible samples, but rather is a model-based measure of deviation from the population value. Ipsos, for example has proposed the credibility interval (Ipsos, 2012) for their estimates from an opt-in panel survey. As noted in Section 6, the credibility interval is measure of uncertainty that is used with Bayesian methods, and Ipsos described their procedure as Bayesian. Other model-based approaches also produce estimates of precision such as standard errors that could be used and do not refer to the average over all possible samples (the accepted terminology for design-based inferences used in probability samples).
>
> Although the research base does not exist to endorse this particular measure or to urge its adoption across the industry, we believe the industry needs constructive attempts to develop measures that fills the gap created by the unsuitability of the standard margin of error calculation with non-probability samples. Treating estimates as though they had no error at all is not a reasonable option. At this point, it falls to individual researchers to judge the usefulness of this particular measure. Such judgments are only possible when organizations using them fully disclose the full range of information specified in the AAPOR Code of Professional Ethics and Practice along with a detailed description of how the underlying model was specified, its assumptions validated, and the measure calculated.

The relevant section of AAPOR's Code of Professional Ethics and Practice now reads as follows:

> III. Standards for Disclosure
>
> A. Disclosure Items for Surveys
>
> 10. Sample sizes (by sampling frame if more than one was used) and a discussion of the precision of the findings.
>
> For probability samples, the estimates of sampling error will be reported, and the discussion will state whether or not the reported margins of sampling error or statistical analyses have been adjusted for the design effect due to weighting, clustering, or other factors.
>
> Disclosure requirements for non-probability samples are different because the precision of estimates from such samples is a model-based measure (rather than the average deviation from the population value over all possible samples). Reports of non-probability samples will only provide measures of precision if they are accompanied by a detailed description of how the underlying model was specified, its assumptions validated and the measure(s) calculated. To

> avoid confusion, it is best to avoid using the term “margin of error” or “margin of sampling error” in conjunction with non-probability samples.

AAPOR has issued detailed guidance in *AAPOR Guidance on Reporting Precision for Nonprobability Samples* (a copy has been provided to the Advisory Panel):

- Guidance is reported for Bayesian credible intervals as well as some other approaches.
- The guidance also includes an example of a statement that could be included in a survey report.
- The document notes that use of one these estimates for precision for a non-probability sample is <u>not</u> required:

  > For some surveys (e.g., exploratory, internal research) estimating precision may not be important to the research goals. For other surveys precision measures may be relevant, but the researcher may not have the statistical resources to compute them. Under the AAPOR Code, it is acceptable for researchers working with nonprobability samples to decline to report an estimate of variance. In such cases, it may be useful to note that the survey estimators have variance, but there has been no attempt to quantify the size.

Note the current 4.3.3 standard can be interpreted as not permitting alternative methods of statistical inference such as Bayesian credible intervals. This is based on the sentence in 4.3.3 #3, *Any use of descriptive statistics must clearly indicate that they are not <u>formally</u> generalizable to any group other than the sample studies, <u>and there cannot be any formal statistical inferences about how the descriptive statistics for the sample represent any larger population</u>.*

Note that the MRIA guidelines does not forbid using Bayesian credible intervals, as the language used refers only to margin of sampling error. Bayesian credible intervals are not a margin of sampling error. The MRIA guidelines are silent on the use of Bayesian credible intervals.

The objective is to update and clarify section 4.3.3 on use of statistical measures with non-probability samples.

Options for revising Section 4.3.3: *Statistical Treatment of Non-probability Samples*

Note that 4.3.3 should be reviewed in the context of the discussion of 4.1 and 4.3.2, where revisions associated with non-probability sampling were discussed.

There are two alternatives for revising 4.3.3:

- ***Option 1:*** **Expressly forbid the use of alternative measures of precision such as Bayesian credible intervals in GC POR surveys.**

  Rationales could be: (1) There is currently not enough evidence that these adequately address concerns about the risk of bias in non-probability samples; (2) There is currently not enough evidence as to what modeling assumptions are appropriate for Bayesian credible intervals in the context of likely sources of non-probability samples for GC POR such as the opt-in panels available in Canada; (3) evaluation of the models requires statistical expertise that users of survey results may not have, which creates a risk of accepting models that are inadequate.

- ***Option 2:*** **Expressly permit the use of alternative measures of precision such as Bayesian credible intervals in GC POR surveys.**

  The rationale could be that alternatives such as Bayesian credible intervals are well-developed and rigorous statistical methods, even if application to public opinion surveys is limited to date. Given (a) the cost and targeting advantages of using opt-in panels, and (b) the imperfections of telephone RDD samples (e.g. low response rates), development and use of these alternatives should be supported.

Proposed Option 2 revision to Section 4.3.3: *Statistical Treatment of Non-probability Samples*

The detailed proposal below is written for *Option 2*. Note, though, that some of the revisions would apply equally to *Option 1*. The main point of difference is in 4.3.3 #3.

4.3.3. Statistical Treatment of Non-probability Samples

1) There can be no statements made about margins of sampling error on population estimates when non-probability samples are used.

   Also, there can be no statement that the sample has a level of error equivalent to that of a probability sample of similar size.

   There can be no tests of statistical significance that are based on estimates of sampling error. An exception to this is a non-probability survey that employs an experimental design in which respondents are randomly assigned to different cells. In this case it is appropriate to use statistical significance tests based on estimates of sampling error to compare results from different cells.

2) The survey report must contain a statement on why no margin of sampling error is reported, based on the following template: "Respondents for this survey were selected using non-probability sampling methods. Because of this, margins of sampling error and tests of statistical significance based on sampling error cannot be reported."

   This statement must be prominently placed in descriptions of the methodology in the survey report, including the executive summary.

3) There are alternatives for margin of sampling error for estimating precision that can be used with non-probability samples, such as Bayesian credible intervals. Researchers have a choice of whether or not to use these alternatives.
   - For some surveys (e.g., exploratory, internal research) estimating precision may not be important to the research goals. For other surveys precision measures may be relevant, but the researcher may not have the statistical resources to compute them. It is acceptable for researchers working with non-probability samples to decline to report an estimate of precision. In such cases, the report must note that survey estimators have variance, but there has been no attempt to quantify the size.
   - If an alternative statistical measure of precision such as Bayesian credible intervals is used:
     - OPTIONAL: The following alternatives are accepted:
       - Bayesian credible intervals
       - Resampling approaches
       - Taylor series linearization
     - The statistical measure of precision that will be used must be stated in the research proposal, together with a rationale and brief description.
     - The survey report must provide:

- A detailed description of how the underlying model was specified, its assumptions validated and the measure(s) calculated. Refer to the AAPOR document *AAPOR Guidance on Reporting Precision for Nonprobability Samples* for the information to provide, as well as an example of the type of statement to make in the report.
- One key assumption is that the survey results are unbiased. This assumption must be prominently noted, together with any limitations on this assumption (see 4.3.1).
- An explanation of how to understand the measure of precision.
- If applicable, how tests of statistical significance of differences or relationships based on the alternative method are to be understood.

Questions Posted on the Discussion Board

C1. The Background lays out two options for the role of statistical methods of reporting precision for non-probability public opinion research surveys for the Government of Canada:

- Option 1: Expressly forbid the use of alternative measures of precision such as Bayesian credible intervals in GC POR surveys
- Option 2: Expressly permit the use of alternative measures of precision such as Bayesian credible intervals in GC POR surveys

Which option do you prefer? Or is there a third option?

C2. Under the Option 2 scenario, the Background proposes a revision to 4.3.3, *Statistical treatment of non-probability samples*. Some of the revisions are specific to Option 2, while others would apply under either Option 1 or Option 2. Regardless of whether or not you favour Option 2:

- Are there any changes to make to the proposed revision to 4.3.3?
- Should the part labeled "OPTIONAL" be deleted, or retained/modified? PORD tilts toward deleting it because it is too specific/prescriptive, but would like the views of the Panel.

D. Statistical Treatment of Census Survey Results

D1. Possible consistency issue in Section 4.6 *Census Surveys*

PORD has received some industry feedback that part of what the Standards say about the statistical treatment of census survey results may not be correct.

The relevant language is in Section 4.6 *Census Surveys*. The following is an abridged version of 4.6:

4.6. Census Surveys

In a census survey, an attempt is made to collect data from every member of a population. For example, an organization might want to do a survey of all of its employees. In this case, the population is "all of the organization's employees", and this would qualify as a census survey if all employees are invited to participate in the survey.

The list whereby all members of the target population are to be contacted and invited to respond must be clearly described, including any of its limitations/exclusions in representing that target

population. Whenever possible, an estimate of the percentage of the population that is excluded from the list must be provided and the potential impact of their exclusion on the research results considered. …

1) The number of attempted re-contacts and procedure for attempted re-contact must be stated.

2) Do not state a margin of sampling error, as this does not apply to census surveys because no sample is drawn.

   The survey report must contain a statement on why no margin of sampling error is reported, based on the following template: "Since the entire population of [target population] was invited to participate in this study there is no margin of sampling error to be estimated or reported. The potential impact of non-sampling error due to non-response is discussed in the results section of the report. [If weighting was done, state the following sentence on weighting:] The data have been weighted to reflect the composition of [the target population (if known) or the sampling frame (e.g., client-supplied list)] on the main known characteristics."

3) There is no need to use inferential statistical tests since the results (frequencies, percentages) reported in a census survey describe the entire target population. Statement #1

   However, it is acceptable to use statistical significance tests to measure differences between sub-groups within the target population. Statement #2

   As with any surveys, be they sample or census, the impact on the results of non- sampling error due to non-response must be assessed to the extent possible, and appropriate caveats on the interpretation of the results must be clearly stated.

The feedback was that Statement #2 is not consistent with Statement #1: *"Using the same reasoning, a sub-group of the census would still be a census of the sub-group. Statistical significance tests would only be appropriate on random samples of the census survey."*

Questions Posted on the Discussion Board

D1.1 Are the two statements in Section 4.6, *Census surveys*, about statistical treatment of results inconsistent? That is:

- Statement #1: *There is no need to use inferential statistical tests since the results (frequencies, percentages) reported in a census survey describe the entire target population.*
- Statement #2: *However, it is acceptable to use statistical significance tests to measure differences between sub-groups within the target population.*

If inconsistent, or not quite right, what changes do you suggest to the language in 4.6, *Census surveys*?

D2. Does response rate affect whether or not a survey is a "census"?

Questions Posted on the Discussion Board

D2.1 In a census, an attempt is made to contact every individual in the population (subject to coverage error). However, rarely will an interview be completed with every attempted contact – that is, the response rate will usually be less than 100%.

Does the survey cease being a census/attempted census if response rate falls below a certain level?

If so, what is the numeric response rate threshold to put in the Standards? And if the response rate is below the threshold, does "margin of sampling error" come into play in terms of calculating confidence intervals and significance tests? Or, is some other measure of statistical precision used?

E. Statistical Treatment of Probability Survey Results

PORD would like the Panel to consider the points made in a short article by Robert Peterson with the provocative title, *It's time for pollsters to report margins of error more honestly*. A copy has been sent to Panel members.

The article lists the following problems with current practices:

> First, the reported margin of sampling error may imply that it pertains to an entire poll. In reality, it is only for a single question, even though research shows that many of us think a reported margin pertains to the entire poll.
>
> Second, the single question used to calculate the margin is a fictitious one. It does not reflect the actual response to any question posed in the poll.
>
> And finally, the formula for calculating the margin arbitrarily assumes that the answer to the fictitious question is 50 percent "Yes" and 50 percent "No." Only if, by chance, were the polled individuals to respond exactly that way would the reported margin be correct for that question.
>
> The rationale offered for reporting margins of error this way is that a 50/50 split has the largest possible margin. If a poll reports the maximum sampling error possible, it is covering the worst-case scenario — for that (fictitious) question.
>
> Except that the poll hasn't necessarily done so. If a question has more than two possible answers — like, "Don't Know" in addition to "Yes" and "No" — it requires a different formula. For subgroups such as "males" or "65 and over," the margins will be larger than that for the overall poll.
>
> Most important, margins of error can be cumulative. Many answers to questions in a poll are related to one another. Collectively, their margins add up to more than that for a single question. In a simple poll with only four questions, and a 95 percent confidence rate for each one, the total margin of sampling error of the poll might be as large as 19 percent.

The article lists the following types of solutions to these problems:

> There are more accurate ways for polls to represent their margins of error. One option is to present none at all. A recent Harvard CAPS-Harris Poll did just that.
>
> Another option is to compute the individual sampling error margins for every question in a poll. Pollsters could then work out an average margin across all poll questions. Such estimates would still possess weaknesses, but they would still be superior to the single fictitious margin currently being reported.
>
> Ultimately, pollsters should devise a multiquestion approach. Such an approach would simultaneously take into account the number, types, weights and relationships of questions, along with individual sampling error margins for actual answers and one for the overall poll.

Standard 14.7 (online)/15.7 (telephone) addresses one of the points in the article, namely that margin of sampling error varies depending upon the observed percentage:

> **14.7/15.7. Mandatory Survey Report Requirements; Results**
>
> 2) For probability samples, state the level of precision, including the margin of error and confidence interval for the total sample and any key sub-groups.

Proposed revision to Section 14.7/15.7 #2

The proposed revision to 14.7/15.7 #2 adds a requirement to indicate how the margin of sampling error varies across different observed percentages. It does not address all of the issues raised in Peterson article, nor does it embrace the recommended solutions (as we understand those).

> **Revised 14.7/15.7. Mandatory Survey Report Requirements; Results**
>
> 2) For probability samples, state the level of precision, including the margin of sampling error and confidence interval for the total sample and any key sub-groups, and for a selection of different percentage values spanning the range of percentages that appear in the report.

Questions Posted on the Discussion Board

E1. Do you think the Standards should be revised to address issues raised in the Peterson article, *It's time for pollsters to report margins of error more honestly*? If so, which issues should be better addressed in the Standards? What do you think of the proposed revision to Section 14.7 (online)/15.7 (telephone), which would now require reporting level of precision both for a range of sample sizes and observed percentages?

F. Online sample information to include in Section 1: *Proposal Documentation*

The current Standards contain the following requirements for proposal documentation when an online sample provider is used:

1.2.4. Description of Data Collection

3) For access panels, a description of the following must be provided, at minimum (when multiple panels are used, the information must be provided for each):
   a) active panel size (provide the definition of "active");
   b) panel recruitment;
   c) panel monitoring;
   d) panel maintenance;
   e) panel refreshment.

The objective is to update and revise this standard, including:

- Distinguish between the information requirements for probability-based panels versus non-probability panels
- Recognize that river sampling may be used as a sample source
- Provide clarity as to what specific information is to be included in the proposal
- Provide guidance, if possible, on good practices the client should look for

Note that there are other relevant sections in the Standards, notably:

- Section 4, *Sampling Procedures*
- Section 5.1.5, *Use of respondent lists*
- Section 1.2.4. #5: *Describe the planned fieldwork validation methods and procedures*

The intent is to not duplicate standards in these areas. Therefore, the focus here is specifically on additional proposal documentation requirements when an online sample provider is used who will be providing sample using an access panel or river sampling.

In 2008 *The Advisory Panel on Online Public Opinion Survey Quality* recommended a number of standards and guidelines pertaining to access panels. For the most part, these were not incorporated into the Online Standards. However, PORD has asked that the current Panel consider whether any aspects of the earlier recommendations should be incorporated into the updated Online Standards. The relevant section of the earlier Panel's report is provided.

The following are some references, which we can send upon request:

- ESOMAR, *28 Questions to Help Research Buyers of Online Samples*
- ESOMAR/GBRN, *Guideline for Online Sample Quality*

Proposed revision to Section 1.2.4. #3

Note that the proposed revision below splits the current #3 into two sections – now numbered #3 and #4, in order to address both probability panels and non-probability sources.

1.2.4. Description of Data Collection

3) When the sample is a **probability sample** of the survey's target population drawn from a probability-based panel, a description of the following must be provided, at minimum (when multiple panels are used, information must be provided for each):

   - Active panel size (provide the definition of "active"), and if possible the active panel size corresponding to the survey's target population that is available to be surveyed

     The panel size should be large enough to ensure a high likelihood the target sample size of usable completed questionnaires can be obtained, taking into account such factors as likely response rate, exclusion rules involving past survey participation, and sample reduction due to data cleaning.

   - How the panel is constructed and maintained such that it is a probability sample
   - How a probability sample of the survey's target population will be drawn from the panel
   - Whether or not respondents will be sourced from more than one panel. If more than one panel is being used, provide the following:
     -- The identity of the other panels
     -- Whether or not each of the other panels is a probability-based panel

        If the sample includes respondents from non-probability samples, then the sample as a whole is a non-probability sample.

     -- The reasons for sourcing sample from more than one panel (e.g. to achieve the desired survey sample size; to get better coverage of the survey population)
     -- Steps to be taken to ensure that people enrolled in more than one of the panels are not sent the survey more than once
   - The incentive or reward respondents are offered for taking part in the survey

     This should be reviewed to ensure it does not risk causing certain types of people to either take or avoid the survey, and thereby potentially bias the sample.

4) When the sample is a **non-probability sample** drawn from a panel or other online sources and obtained from an online sample provider, a description of the following must be provided, at minimum (when multiple panels are used, information must be provided for each). Note that "panel" refers both to panels operated by an online sample provider and to lists available from online sample providers.

   - Whether all or part of the sample will come from web-intercept methods ("river sampling") or other non-panel sources, and if so what these sources are and how respondents will be selected for your survey

     These methods may involve less quality control in terms of validation of respondent identity. The nature of the websites from which respondents are sourced can potentially introduce bias: a large number of diverse websites is better than a small number of websites.

   - If a panel is being used, the active panel size (provide the definition of "active"), and if possible the active panel size corresponding to the survey's target population

     The panel size should be large enough to ensure a high likelihood the target sample size of usable completed questionnaires can be obtained, taking into account such factors as likely

response rate, exclusion rules involving past survey participation, and sample reduction due to data cleaning.

- If a panel is being used, whether or not respondents will be sourced from more than one panel. If more than one panel is being used, provide the following:
  - -- The identity of the other panels.
  - -- The reasons for sourcing sample from more than one panel (e.g. to achieve the desired survey sample size; to get better coverage of the survey population; to reduce the risk of bias)
  - -- Steps to be taken to ensure that people enrolled in more than one of the panels are not sent the survey more than once.
- If a panel is being used, describe the recruitment sources used to populate the panel on an ongoing basis, and if possible information on the proportion of panelists accounted for by each source.

  The goal is to ensure the diversity of the survey population is represented, not only in terms of demographics but also on relevant behaviours and attitudes. In this regard, it can be better to have a large number of diverse sources, to reduce the risk of bias.
- If a router is used to select survey respondents, what measures are taken to guard against, or mitigate, any bias that might arise from using a router.

  A router is an online system that screens potential respondents and assigns each person to one of multiple surveys being fielded at the same time. Routers are important for efficiently fielding multiple surveys. However, there can be a risk that systematically directing people with certain characteristics to other surveys may bias the profile of people directed to your survey.
- If profile data from a panel is used to structure the sample or for key analyses, and these measures are not validated by remeasuring the characteristics in the questionnaire, provide information on the panel's validation processes to ensure the information is accurate and up to date.

  Panel profile data on panelists can be very helpful for structuring a sample. However, it is a good practice to verify this data by remeasuring the characteristics in the survey questionnaire.
- The incentive or reward respondents are offered for taking part in the survey

  This should be reviewed to ensure it does not risk causing certain types of people to either take or avoid the survey, and thereby potentially bias the sample.

Questions Posted on the Discussion Board

F1. **Focusing on the proposal related to probability-based panels:** Is it necessary or useful to split Section 1.2.4 #3, *Proposal Documentation, Description of Data Collection,* to have a separate section on probability-based panels? Do you have any suggestions for changes to the proposed section on probability-based panels (i.e. the "new" 1.2.4 #3 in the Background)?

F2. **What is "large enough" for a panel?:** For both probability and non-probability surveys in the revised Section 1.2.4 #3/#4, it says the following about "active panel size": *The panel size should be large enough to ensure a high likelihood the target sample size of usable completed questionnaires can be obtained, taking into account such factors as likely response rate, exclusion rules involving past survey participation, and sample reduction due to data cleaning.*

PORD has asked the Panel for advice on how to assess "large enough."

Alternative listings of requirements for non-probability samples from an online sample provider

As noted at the beginning of this section, the objective is to update and revise this *Proposal Documentation* Standard 1.2.4 #3, with the focus here being on non-probability samples obtained from an online sample provider. The criteria to apply include:

- Recognize that river sampling may be used as a sample source
- Provide clarity as to what specific information is to be included
- Provide guidance, if possible, on good practices the client should look for

To this list of criteria we also add "simplicity": simpler/shorter language is preferred as long as it meets the objectives of being clear and useful.

PORD has asked that the Panel draw upon three different listings of what the required proposal documentation should be, when formulating suggested requirements for non-probability samples obtained from an online sample provider:

- The current *Proposal Documentation* 1.2.4 #3
- ESOMAR *Online Research*, Section 6.1, *Sample Source and Management*
- Proposed revision above in the Background

The following pages show the guidance from the three sources. As you can see, they differ in length/complexity.

As noted earlier, PORD has also asked that the Panel consider the suggestions of the earlier *Advisory Panel on Online Public Opinion Survey Quality*. The relevant excerpt has been sent separately to the Panel.

Questions Posted on the Discussion Board

F3. **Focusing on non-probability samples obtained from an online sample provider:** This is complicated but we ask that you take a crack at it. Please recommend revisions to Section 1.2.4 #3, for non-probability samples – that is, what information to give about the sample in *Proposal Documentation*. The Background provides some alternatives to pick and choose from (current standard, ESOMAR, proposed revision 1.2.4 #4, suggestions from the 2008 *Advisory Panel on Online Public Opinion Survey Quality*). The criteria to apply are:

- Recognize that river sampling may be used as a sample source
- Provide clarity as to what specific information is to be included
- Provide guidance, if possible, on good practices the client should look for
- Simplicity

1.2.4. Description of Data Collection – Alternative approaches

3) For access panels, a description of the following must be provided, at minimum (when multiple panels are used, the information must be provided for each):

Current 1.2.4	**ESOMAR *Online Research* (pp. 18-19)**	**A proposed revision to 1.2.4 #4 (the non-probability section)**
• active panel size (provide the definition of "active"); • panel recruitment; • panel monitoring; • panel maintenance; • panel refreshment.	• details of how the sample was recruited and a description of the sampling frame and how well the sample covers the target population it is meant to represent • steps taken to validate sample sources • the procedures used to "on-board" prospective participants to panels • cleaning and updating procedures • any monitoring of survey-taking performance or quality controls to minimize satisficing or fraud and the steps taken if such behaviour is identified** • whether and how new sources are integrated • any procedures in place to maximize sample consistency for tracking projects	When the sample is a non-probability sample drawn from a panel or other online sources and obtained from an online sample provider, a description of the following must be provided, at minimum (when multiple panels are used, information must be provided for each). Note that "panel" refers both to panels operated by an online sample provider and to lists available from online sample providers. • Whether all or part of the sample will come from web-intercept methods ("river sampling") or other non-panel sources, and if so what these sources are and how respondents will be selected for your survey These methods may involve less quality control in terms of validation of respondent identity. The nature of the websites from which respondents are sourced can potentially introduce bias: a large number of diverse websites is better than a small number of websites. • If a panel is being used, the active panel size (provide the definition of "active"), and if possible the active panel size corresponding to the survey's target population The panel size should be large enough to ensure a high likelihood the target sample size of usable completed questionnaires can be obtained, taking into account such factors as likely response rate, exclusion rules involving past survey participation, and sample reduction due to data cleaning. • If a panel is being used, whether or not respondents will be sourced from more than one panel. If more than one panel is being used, provide the following: -- The identity of the other panels. -- The reasons for sourcing sample from more than one panel (e.g. to achieve the desired survey sample size; to get better coverage of the survey population; to reduce the risk of bias) -- Steps to be taken to ensure that people enrolled in more than one of the panels are not sent the survey more than once.

		• If a panel is being used, describe the recruitment sources used to populate the panel on an ongoing basis, and if possible information on the proportion of panelists accounted for by each source. The goal is to ensure the diversity of the survey population is represented, not only in terms of demographics but also on relevant behaviours and attitudes. In this regard, it can be better to have a large number of diverse sources, to reduce the risk of bias. • If a router is used to select survey respondents, what measures are taken to guard against, or mitigate, any bias that might arise from using a router. A router is an online system that screens potential respondents and assigns each person to one of multiple surveys being fielded at the same time. Routers are important for efficiently fielding multiple surveys. However, there can be a risk that systematically directing people with certain characteristics to other surveys may bias the profile of people directed to your survey. • If profile data from a panel is used to structure the sample or for key analyses, and these measures are not validated by remeasuring the characteristics in the questionnaire, provide information on the panel's validation processes to ensure the information is accurate and up to date. Panel profile data on panelists can be very helpful for structuring a sample. However, it is a good practice to verify this data by remeasuring the characteristics in the survey questionnaire. • If the survey is a tracking study, the steps that will be taken to ensure comparability of the sample with previous waves of the tracking study • The incentive or reward respondents are offered for taking part in the survey This should be reviewed to ensure it does not risk causing certain types of people to either take or avoid the survey, and thereby potentially bias the sample.

** Section 7.8 of the Standards addresses satisficing. The question is whether this should also be referenced in Section 1, *Proposal Documentation*

7.8. Detecting and Dealing with Satisficing

1) The access panel provider, working with the client, shall implement procedures to identify and remove fraudulent and inattentive panel members or respondents, documenting these procedures and actions taken.

G1. Required Questions: Introduction wording

Section 2.1 #3 of the Online and Telephone Standards gives required questions that must be asked in all surveys of individuals, *"unless a convincing argument is made that the research objectives are better served by excluding one or more of them"* (there is also an exclusion for B-to-B research where the unit of analysis is the organization).

Section 2.1 #3 states:

> The wording used for each question must be that provided below, unless a convincing argument is made that particular research objectives require alternative wording. Even in these exceptional cases, the terms used and/or categories applied (e.g., for household income) to capture responses must be those provided below.

Section 2.1 #3 states the following rationale for these required questions:

> The data from the age, education, and language questions (along with the recording of geographic location and sex) allows comparison with Statistics Canada census data for the purpose of non-response analysis. The data, along with that from the employment status and income questions, also facilitate the comparison of results between Government of Canada public opinion research studies. (See Section 8. for further detail on non-response bias.)

Note that comparability to Statistics Canada's questions and response options is a critical requirement. This is needed for non-response analyses, and it is needed when the variables are used to weight survey data to match the population.

Because of differences between the online (self-completion) and telephone (interviewer-administered) modes, the questions/response options may be somewhat different for the two modes.

In this regard, an issue for the Panel to keep in mind throughout Section F when there are differences in wording between the online and telephone modes is what constitutes comparability with Statistics Canada's questions and response options. A key example: we can access the self-completion version of the 2016 Census questionnaire, but it is not always clear what Statistics Canada would consider to be a comparable question in a telephone survey.

Introduction Wording

There is a general requirement (Section 2.1 #2) to inform respondents at the beginning of a survey of the confidentiality of their questionnaire responses, but the current Standards do not state any specific wording for how to preface asking the required demographic questions – which are often asked near the end of the survey.

For reference, the relevant parts of Section 2.1 #2 are:

> 2.1. #2 The following are required elements of all Government of Canada online survey questionnaire introductions:
>
> f) inform respondents that their participation in the study is voluntary and completely confidential;

g) inform respondents that their responses remain anonymous. In the exceptional cases where the research objectives require that respondent identity be revealed, the informed consent of the respondent must be obtained;

The Privacy Commissioner has requested the addition of a statement on privacy before the demographics section of questionnaires, such as the one below.

> *These last few questions are for statistical purposes and will be kept confidential. Your identity will always remain anonymous.*

Proposed addition to Section 2.1

The proposal is to insert a new standard just before the current 2.1 #3 as follows (the current 2.1 #3 would become 2.1 #4:

2.1. #3 The following statement is required for all Government of Canada telephone/online surveys prior to administering the demographic section of the questionnaire:

> *These last few questions are for statistical purposes only. Your answers will be kept anonymous and confidential and will be combined with the answers from other respondents to this survey.*

Questions Posted on the Discussion Board

G1.1 The first sentence in the proposed addition of a new Section 2.1 #3 is, *"These last few questions are for statistical purposes only."* Is "statistical purposes" the best phrase to use, or is there a different phrase that would be more clear/effective? For example, "analytical purposes", or "purposes of analysis"?

G1.2 Do you have any suggestions for revising the proposed Section 2.1 #3?

G2. Required Questions: Gender

The following is the current mandated question in Section 2.1 #3 for telephone and online surveys:

Telephone Surveys	***Online Surveys***
Gender: [Do not ask: record based on interviewer observation]	**Gender**: What is your gender? Female Male Prefer not to answer

Note: The fact that the answer options are different by methodology may be problematic for mixed-mode surveys as well as for weighting -- i.e. telephone is based on the interviewer's opinion and gives two answer options, while the online is based on respondent self-classification and gives three answer options. It is also an issue in its own right if there is an interest or a need to compare different surveys using the two different methods.

There are several issues related to the current 'gender' question for the Panel to consider.

- Demographic data is collected for analytical purposes, to assess the representativeness of the sample, and for use in weighting the data (if necessary).
- Also, as stated in the Standard, there is a need to allow for comparison of data collected in Government of Canada public opinion research (POR) surveys to a high quality data source, e.g. Statistics Canada census data for the purposes of non-response bias analysis or weighting. To perform these types of analyses, the questions must be asked in the same way in POR surveys as they are on the high quality data source.
- Statistics Canada's 2016 census has the following question (used in the paper/online questionnaire):

  What is this person's sex?
  Male
  Female

  The census questionnaire in other major countries (e.g. U.S., United Kingdom, Australia) also ask a similar question and offer only the two answer options shown above on their most recent census data collection cycle.
- We're not sure how Statistics Canada collects data on sex in a census telephone survey – i.e. is it based on interviewer observation, or is a question asked – and if so what is the question?
- There is a difference between a person's sex and a person's gender. While they are related, they are not the same thing. Sex is binary and refers to one's biology at birth whereas gender is not binary can often be about how one identifies oneself physically and/or psychologically.
- The census bureaus in all the major countries noted above have committed to including a question(s) to accurately capture 'gender' identity for the 2021 census and all have started secondary and primary research to test questions in preparation for their next census.
- There are various organizations (e.g. Equality and Human Rights Commission, Rainbow Health Ontario, the Williams Institute) and a private sector firm – GfK Research – that have looked at how best to ask these questions but there is no general consensus on either how best to ask the questions or what answer options should be provided to accurately capture the information.
- In January 2016, the U.K. market research standards organizations MRS published *MRS Guidance on Collecting Data on Sex and Gender*. In their guidance they state:

  > It is important that all research participants feel satisfied that their voice is being heard. Allowing participants to provide information in a way that reflects the view they want to express is explicitly set out in the MRS Code of Conduct and this also applies to describing their sex or gender. The categories provided for completion must facilitate and recognise the fluidity of gender identity, by providing an additional option such as free-field 'other'.
  >
  > Members must ensure that participants will be able to proceed through any research without being required to indicate male or female options if they do not identify as such. This may be achieved by including female, male and other as (free field) options; allowing participants to self-identify by asking an open-ended question or by adopting a two-step approach that covers assigned sex at birth and then current gender identity.

  MRS also notes the following key points:
  - ✓ Collection of sex and/or gender information needs to be tailored to the data collection tool.

- ✓ Self-completion tools should at a minimum include an additional field of "other" for either gender or sex questions.
- ✓ Interviewer-aided techniques for collecting information need to build in a level of discretion and flexibility for interviewers, to ensure questions are responsive and are handled appropriately.

MRS provided the following examples of questions and a checklist of what questions researchers should ask themselves before deciding what question to ask:

What is your sex?	*What is your sex? OR What is your gender?*
Male *Female* *Intersex* *Prefer not to say*	*Male* *Female* *Other (please specify)* *Prefer not to say*

Checklist - Questions to ask

Do I need to collect information on either sex or gender? Is there a research purpose for collecting this information?

- o Why?
- o Is it relevant and not excessive?

Should I collect information on sex or should I collect gender?

- o Do I want to match with other data sets or sources? If so, what do they collect?

Can I make the answers optional?

What response options should I provide?

- o Should I provide closed categories (including other) or an open field?
- o Do I want to include an intersex option for sex or an "other" option?

The MRS goes on to say the following about gender identity questions specifically:

> Establishing best practice in developing and asking gender identity questions will need to build on the position and practice of the ONS [*Office of National Statistics* which is the UK equivalent of Statistics Canada] and research carried out by the EHRC [*Equality and Human Rights Commission*]. This will allow the research community to design and implement a consistent and standard gender identity question that can be understood and answered by all people living in the United Kingdom (UK).

- Uptal Dholakia in his article *How Should Market Researchers Ask About Gender in Surveys*? in an online blog for Psychology Today (September 2016) points out that:

> Any well-designed market research survey is based on two core principles: the ***principle of accuracy*** and the ***principle of inclusiveness***. A questionnaire should be designed to gather information accurately, using best practices of survey design that psychometricians have formulated over several decades. But this is not enough. A survey should also be inclusive. When a respondent has finished taking a survey, they should feel like the opinion they have provided will be valued just as much as every other survey taker.

Proposed revision to Section 2.1. #3 Gender

We propose the Advisory Panel consider two alternative options as a starting point, which the Panel can of course revise as they see fit:

- ***Option 1: Status quo*** – Leave the standards unchanged

The status quo is to ask about "gender", as follows:

Telephone: Binary as recorded by the interviewer (Male, Female)
Online: Three response options (Male, Female, Prefer not to answer)

The advantage of Option 1 is that it is more inclusive in the online version, where the question explicitly allows an individual not to answer the question.The disadvantages of Option 1 include:

-- The online version of the question does not match the Statistics Canada question in two respects: it uses "gender" rather than "sex", and it offers three response options rather than two. As a result, from the perspective of using the results for assessing sample representativeness and for weighting, it is less comparable to Statistics Canada census data.

-- A fundamental demographic question has a different minimum number of response options depending on survey mode. This would not be a problem if the larger number of response options (3) could be logically collapsed down to the smaller number of response options (2). However, "prefer not to answer" cannot be collapsed into either "male" or "female", nor is it semantically accurate to divide it between these two categories. The telephone version does not give the interviewer an option to record either "don't know" or "refuse/prefer not to say."

- ***Option 2: More closely match Statistics Canada question(s)***

The advantages of Option 2 (see proposed text below) are:

-- A major reason the Standards require asking the question is to ensure that accurate population statistics are used for assessing sample representativeness and for weighting. It follows directly from this that the question should be asked in the same way as it is asked by Statistics Canada (the usual source of population statistics).

The telephone question is similar to but does not exactly match Statistics Canada in terms of giving the interviewer a main choice of two response options – but there is a possibility of recording a third option of "prefer not to answer."

However, the online version does not exactly match the Statistics Canada question in either the question (they ask "sex" rather than "gender"), or the number of response options. This makes the sex/gender data less comparable for purposes of measuring sample representativeness and weighting.

-- Using the proposal below, the number of response options is the same (2) regardless of survey mode.

Neither the telephone nor the online question exactly match the Statistics Canada question because of the possibility of a "prefer not to answer" response.

Proposed *Option 2* revision to the Standard

Telephone Surveys	*Online Surveys*
Sex: [Do not ask: record based on interviewer observation] If uncertain based on voice, ask: What is your sex? [read list] Female Male Or do you prefer not to answer	**Sex**: What is your sex? Female Male Prefer not to answer

Questions Posted on the Discussion Board

G2.1 With respect to the required "gender" question in Section 2 of the Standards, the Background gives two options:

- Option 1: Status quo
- Option 2: Revise the online and telephone questions

Which option do you favour?

G2.2 With respect to Option 2, what do you suggest be the telephone and online questions. Do you agree with the proposal in the Background, or do you suggest something else? Do you think an "other (please specify)" category should be added as in the example MRS question?

G3. Required Questions: Language

The following are the two mandated questions for language for telephone and online surveys in Section 2.1 #3. The researcher can choose to use both questions or one of the two questions, depending on the survey objectives:

Telephone Surveys	*Online Surveys*
Mother Tongue: What is the language you first learned at home as a child and still understand? [READ LIST — ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] [DO NOT READ] Don't know/Refused	**Mother Tongue:** What is the language you first learned at home as a child and still understand? [ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] Prefer not to answer
Language Spoken at Home: What language do you speak most often at home? [READ LIST — ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] [DO NOT READ] Don't know/Refused	**Language Spoken at Home:** What language do you speak most often at home? [ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] Prefer not to answer

PORD has requested input on what question or questions to ask to best get at official language of respondents.

Considerations include:

- In most surveys, there are at least two sets of data input on language:
  1) Currently, given that it is a requirement to ask and offer respondents the choice of English or French to answer a survey, and on the assumption that this is captured as part of the data set for each respondent, information on preferred language for survey completion is already being collected on all surveys.
  2) At least one of the two questions above are required to be included in a survey.
- Also, as stated in the Standard, there is a need to allow for comparison of data collected in Government of Canada public opinion research (POR) surveys to a high quality data source, e.g. Statistics Canada census data for the purpose of non-response bias analysis or weighting. To perform these types of analyses, the questions must be asked in the same way in POR surveys as they are on the high quality data source.
- Statistics Canada census 2016, asks the following main questions (used in the paper/online questionnaire):

  Q.7: *Can this person speak English or French well enough to conduct a conversation?*
  English only
  French only
  Both English and French
  Neither English or French

  Q.8a: *What language does this person speak **most often** at home?*
  English
  French
  Other language -- specify

  Q.9: *What is the language that this person **first learned** at home **in childhood** and **still understands**?*
  English
  French
  Other language -- specify
- Note that the language in which the survey is completed is not equivalent to Q.7 in the census. Also, our view is that "language spoken most often at home" (Q.8a) is more useful as a fundamental measure of language use than is Q.7, so we do not suggest making a question like Q.7 mandatory in surveys.
- There was some input from MRIA members. As we understand it, the general view expressed was that the question *language spoken most often at home* is considered to be the better indicator of official language, and mother tongue was seen often to be less relevant. This is not to say that asking both questions would never have value, but rather is a point of view on which of the two existing questions is more often likely to be useful.

  In the MRIA response to questions from PORD, it was pointed out that when the two existing mandated questions for official language are asked, they *"are the two most closely related to ethnicity."* However, these are imperfect measures of ethnicity and if the study objectives

require collecting data on ethnicity, then there are better questions that get at ethnic origin than inferred from language.

Proposed revision to Section 2.1.3 Official language

The following are three options for the Panel to consider:

- ***Option 1: Status quo*** – Leave the two official language questions unchanged in terms of both the actual question and the option to use one or the other or both questions for official language.
- ***Option 2: Status quo but modified mother tongue question and answer options for both questions*** – Leave the two official language questions unchanged in terms of the option to use one or the other or both questions for official language. However, revise (a) the question for *mother tongue* to match the Stats Can 2016 census questionnaire, and (b) synchronise answer options for both telephone and online for both questions.

Telephone Surveys	***Online Surveys***
Mother Tongue: What is the language you first learned at home in childhood and still understand? [READ LIST — ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] Or do you prefer not to answer	**Mother Tongue:** What is the language you first learned at home in childhood and still understand? [ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] Prefer not to answer
Language Spoken at Home: What language do you speak most often at home? [READ LIST — ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] Or do you prefer not to answer	**Language Spoken at Home:** What language do you speak most often at home? [ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] Prefer not to answer

- ***Option 3: Make language spoken at home the only mandatory question and remove mother tongue question from the standards***.

Telephone Surveys	***Online Surveys***
Language Spoken at Home: What language do you speak most often at home? [READ LIST — ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] Or do you prefer not to answer	**Language Spoken at Home:** What language do you speak most often at home? [ACCEPT ALL THAT APPLY] English French Other [SPECIFY OR DO NOT SPECIFY DEPENDING ON THE NEEDS OF THE STUDY] Prefer not to answer

Questions Posted on the Discussion Board

G3.1 With respect to the required language questions in Section 2 of the Standards, the Background gives three options:

- Option 1: Status quo – Mother tongue and Language spoken at home
- Option 2: Similar to the status quo, but revise the questions to more closely match the Census questions and to synchronize telephone and online answer options
- Option 3: Make language spoken at home the only mandatory question and remove mother tongue question from the standards.

Overall, which option do you favour?

G3.2 Do you have any suggestions for changes to the proposed question wording/response options for Option 2 or Option 3? Or for an "Option 4"?

G4. Required Questions: Age

The current required age question is:

Telephone Surveys	***Online Surveys***
In what year were you born? [Record year – XXXX] [IF PREFERS NOT TO PROVIDE A PRECISE BIRTH YEAR, ASK:] Would you be willing to tell me in which of the following age categories you belong? 18 to 34 35 to 49 50 to 54 55 to 64 OR 65 or older? [DO NOT READ] Refused	In what year were you born? [YYYY] Prefer not to answer [IF PREFERS NOT TO PROVIDE A PRECISE BIRTH YEAR, ASK:] Would you be willing to indicate in which of the following age categories you belong? 18 to 34 35 to 49 50 to 54 55 to 64 65 or older Prefer not to answer

Note that we assume a survey could use more detailed breaks as long as these can be collapsed into the categories above (although the Standards do not explicitly say this).

In the census, Statistics Canada determines age from birth date.

For reference, the 2016 Census Profile for Canada shows the following age distribution:

		% of Population 18+		
0-17	20.0%			
18-24	8.8%	10.9%	18-24	10.9%
25-29	6.5%	8.1%	25-34	16.4%

30-34	6.6%	8.3%		
35-39	6.5%	8.1%	35-44	16.2%
40-44	6.4%	8.0%		
45-49	6.7%	8.4%	45-54	17.9%
50-54	7.6%	9.5%		
55-59	7.5%	9.3%	55-64	17.5%
60-64	6.5%	8.1%		
65-69	5.6%	7.0%	65 or over	21.1%
70-74	4.0%	5.1%		
75-79	2.9%	3.6%		
80 or older	4.3%	5.4%		

Questions Posted on the Discussion Board

G4.1 Should the age categories be left the way they are in Section 2.1 #3, or do you suggest revising them? Revisions could include revisions to the break points, and revisions to the total number of age categories.

PORD is particularly interested in whether a smaller number of categories could be used, and what those categories would be – what would you suggest a smaller number of categories consist of?

G5. Required Questions: Education

The current required education question is:

Telephone Surveys	***Online Surveys***
What is the highest level of formal education that you have completed? [READ LIST]	What is the highest level of formal education that you have completed?
Grade 8 or less	Grade 8 or less
Some high school	Some high school
High School diploma or equivalent	High School diploma or equivalent
Registered Apprenticeship or other trades certificate or diploma	Registered Apprenticeship or other trades certificate or diploma
College, CEGEP or other non-university certificate or diploma	College, CEGEP or other non-university certificate or diploma
University certificate or diploma below bachelor's level	University certificate or diploma below bachelor's level
Bachelor's degree	Bachelor's degree
Post graduate degree above bachelor's level	Post graduate degree above bachelor's level
[DO NOT READ] Prefer not to answer	Prefer not to answer

For reference, the 2016 Census Profile for Canada shows the following education categories:

No certificate; diploma or degree	18.3%
Secondary (high) school diploma or equivalency certificate	26.5%
Postsecondary certificate; diploma or degree	55.3%
Apprenticeship or trades certificate or diploma	9.8%
College; CEGEP or other non-university certificate or diploma	19.4%
University certificate or diploma below bachelor level	2.8%
University certificate; diploma or degree at bachelor level or above	23.3%
Bachelor's degree	15.5%
University certificate or diploma above bachelor level	1.6%
Degree in medicine; dentistry; veterinary medicine or optometry	0.7%
Master's degree	4.6%
Earned doctorate	0.8%

It has been pointed out that the first two categories in the current required education question do not line up well with the Quebec education system. PORD provided the following from Wikipedia:

> *Mandatory elementary education (école primaire) starts with grade 1, through to grade 6. Secondary school (école secondaire) has five grades, called secondary I-V (Sec I-V for short) or simply grades 7-11. Students are 12 to 16 years old (age of September 30), unless they repeat a grade. Upon completion of grade 11, students receive their high school diploma from the provincial government.*

So, in Quebec, Grade 8 is the start of high school (école secondaire), whereas outside Quebec Grade 9 is the start of high school. The view is that the first two existing required categories – *Grade 8 or less*, and *Some high school* – are confusing in Quebec, and indeed the two categories overlap, since Grade 8 is also "some high school" in Quebec.

Proposed revision of the education question

One approach is simply to drop any attempt in the required question to distinguish subcategories of less than a high school diploma (such distinctions can always be added on an ad hoc basis for a particular survey).

In this scheme, the first two response options could be:

Less than a high school diploma or equivalent
High School diploma or equivalent

Questions Posted on the Discussion Board

G5.1 In the context of the differences between the education systems between Quebec and other parts of Canada, are the first two required education response options in Section 2.1 fine as is, or do they need to be revised?

If revised, what do you suggest the response options should be? What do you think of the proposal to eliminate having subcategories of "less than high school"?

G6. Required Questions: Household Income

The current required household income question is:

Telephone Surveys	*Online Surveys*
Which of the following categories best describes your total household income? That is, the total income of all persons in your household combined, before taxes [READ LIST]? Under $20,000 $20,000 to just under $40,000 $40,000 to just under $60,000 $60,000 to just under $80,000 $80,000 to just under $100,000 $100,000 to just under $150,000 $150,000 and above [DO NOT READ] Refused	Which of the following categories best describes your total household income? That is, the total income of all persons in your household combined, before taxes? Under $20,000 $20,000 to just under $40,000 $40,000 to just under $60,000 $60,000 to just under $80,000 $80,000 to just under $100,000 $100,000 to just under $150,000 $150,000 and above Prefer not to answer

Note that we assume a survey could use more detailed breaks as long as these can be collapsed into the categories above (although the Standards do not explicitly say this).

Statistics Canada now sources income data from administrative records, so there is no Statistics Canada question to use as a model.

Statistics Canada reports various income measures, one of which is "total income for private households before tax", which appears to be equivalent to "total household income" in the questions above.

For reference, based on 2015 data, the Census Profile for Canada shows the following for total income for private households:

Statistics Canada – 2015 data – Total income of private HH			
Census Profile for Canada		***Current Required Breaks***	
Under $5;000	1.6%	Under $20,000	9.7%
$5;000 to $9;999	1.4%		
$10;000 to $14;999	2.7%		
$15;000 to $19;999	4.0%		
$20;000 to $24;999	4.3%	$20,000 to just under $40,000	16.7%
$25;000 to $29;999	3.8%		
$30;000 to $34;999	4.3%		

$35;000 to $39;999	4.3%		
$40;000 to $44;999	4.2%	$40,000 to just under $60,000	16.1%
$45;000 to $49;999	4.1%		
$50;000 to $59;999	7.8%		
$60;000 to $69;999	7.2%	$60,000 to just under $80,000	13.7%
$70;000 to $79;999	6.6%		
$80;000 to $89;999	5.9%	$80,000 to just under $100,000	11.2%
$90;000 to $99;999	5.3%		
$100;000 to $124;999	10.4%	$100,000 to just under $150,000	17.7%
$125;000 to $149;999	7.2%		
$150;000 to $199;999	7.9%	$150,000 and above	14.7%
$200;000 and over	6.8%		

Some other statistics using 2015 data:

- Median total income for private HH: $70,336
- Average total income for private HH: $92,764
- Percentage of Canadians living in low-income (LIM-AT[35]): 14.2%

  The dollar threshold for low income varies with household size. Statistics Canada has the following table showing the low-income thresholds as a function of household size[36]:

Low-income measures thresholds		
HH Size	**After-tax income**	**Before-tax income**
1 person	$22,133	$25,516
2 persons	$31,301	$36,084
3 persons	$38,335	$44,194
4 persons	$44,266	$51,031
5 persons	$49,491	$57,054
6 persons	$54,215	$62,500
7 persons	$58,558	$67,508

Questions Posted on the Discussion Board

G6.1 Should the household income categories be left the way they are in Section 2.1 #3, or do you suggest revising them? Revisions could include revisions to the break points, and revisions to the total number of income categories.

[35] LIM-AT is a low-income measure based on after-tax household income

[36] Statistics Canada, Table 4.2, Low-income measures thresholds (LIM-AT and LIM-BT) for private households of Canada, 2015 http://www12.statcan.gc.ca/census-recensement/2016/ref/dict/tab/t4_2-eng.cfm

G7. Possible Additional Required Question: Type of Home Phone(s)

PORD consulted with MRIA as to the value of adding a question in telephone surveys about type(s) of phone(s) a respondent has access to at home:

PORD question to MRIA:

Consider questions on landline vs. cell for classification purposes (landline only; cell only; both) e.g., adding a question at the beginning of the survey similar to the question to cell phone respondents: *At home, do you have a cell phone as well as a traditional telephone line?*

MRIA comments and response:

This question is valid in that it may help the researcher to know if there is a segment of the population in question that was missed. The ratio of landline contacted to cell phone contacted based on the known demographics of landlines and cell phone lines in the area in question may provide additional insights in the analysis.

The information about phone usage may also be incorporated into the weighting scheme for telephone surveys. The literature review commissioned by PORD notes this possibility (see pp. 13-14 of the review). The review says that, *"At the time of writing, there is no consensus on the best approach to weighting dual-frame survey samples."* It goes on to cite an example of an approach that weights by telephone status (cell phone only, landline only, dual phone users), and an approach that does not.

Classification of household telephone status can be quite complicated, depending on the approach taken to weighting, or to how respondents are to be selected for the interview within a household[37]. Some examples of complicating factors:

- People in a household with a landling may nonetheless most often use a cell phone ("cell phone mostly"), and similarly someone with a cell phone may mostly use a landline ("landline mostly").
- An individual may have more than one cell phone.
- People within a household may share a cell phone.
- A cell phone might be personal use only, business use only, or a combination of personal and business use.

The proposal below is to keep things simple and to classify the household telephone status of respondents in a survey of individuals (not surveys of businesses/organizations) into three basic categories: landline only, cell phone only, or dual users (household has both cell phone users and has a landline). Note that this is a classification of the household, not of the individual respondent. It does not attempt to differentiate personal versus business use of a cell phone.

[37] A good overview of the complications of weighting telephone surveys by telephone status can be found in 2010 AAPOR Cell Phone Task Force report, *New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers* – see the section on Weighting. The report also has an appendix showing the questions some major survey organizations were using at the time to ascertain telephone status. The report concluded that there was no consensus on what questions to ask to ascertain telephone status.

A terminology issue to consider is how to refer to a landline phone. Does everyone know what "landline" refers to? If a household has a VOIP phone and no telco landline, would they say they have a "landline"? Do people know what "VOIP", or "IP", phone is? Would people interpret "regular phone" to mean anything other than a cell phone?

The following is a possible question that could be added to Section 2.1 #3

> **Type of Home Phone(s):** The intent of these two questions is to classify the respondent's household into one of three mutually exclusive categories: cell phone only, landline only, dual cell phone and landline.
>
> a) Do you or anyone in your household have a cell phone? Yes/No
> b) Is there a phone inside your home that is currently working and is not a cell phone? Yes/No

Questions Posted on the Discussion Board

G7.1 Should a question on type of home phone service be added as a requirement for all telephone surveys of individuals in Section 2.1 #3?

G7.2 Do you have suggestions for any revisions to the suggested approach to determining type of household phone service?

Appendix B: Background and Questions – Discussion Board #2

DISCUSSION BOARD #2: BACKGROUND AND QUESTION GUIDE
(Start date: May 14; End date: June 1)

Topics:

The following contains background information and the questions for the group discussion. More questions may be developed for this Discussion Board as it progresses and based on the views and suggestions of Panel members.

Please note the following:

1) **You will need to refer to this Background for the detailed commentary on the questions shown on the Discussion Board.** The Discussion Board displays only the questions. This Background also provides, for example:
   -- specific excerpts from relevant sources
   -- references to information available

   This type of detail has **not** been included in the Discussion Board questions.

2) Please answer **all** the relevant questions; it is important to know the point of view of each member of the Advisory Panel.

3) Wherever possible, we need to have as detailed answers as possible, particularly in areas where there is possible disagreement between Panel members on either a principle stated, or the language used to express the principle.

4) We need all members of the panel to begin posting answers to the questions as soon as possible and shortly after the Discussion Board is up and running. Otherwise, there will be little if any time left to debate issues, to try to get clarification or more detail from the Panel, or to come to some agreement on issues where Panel members have different views.

   It is the mandate of the Advisory Panel ***to reach consensus where possible even though it is not an essential outcome of deliberations***. Where there is initial disagreement, please talk to one another to see if a consensus can be found. Think of lack of consensus as a last resort.

5) Just to remind you about the terminology being used:

Standards	Practices which are ***requirements*** for all quantitative research conducted by the Government of Canada
Guidelines	Practices which are ***recommended,*** but would not be requirements; that is, known good practices or criteria that serve as a checklist to ensure quality research but are not necessarily applied to every study

Wherever possible for the sake of simplicity, the standards or guidelines should be the same for Government of Canada (GC) public opinion research (POR) telephone and online surveys.

6) The Public Opinion Research Directorate (PORD) asks that the following criteria be applied when formulating standards and guidelines:

In general, we wish to adhere to the following criteria when revising/developing Standards:

- *Be measurable*
- *Not conflict with other GC departments (e.g., TBS) or related policies or standards (e.g., IT, legal)*
- *Avoid material that will quickly become out-of-date*
- *Not overly prescriptive*
- *Minimize Web links in the content*

7) You will see various citations in the footnotes of this Background. We have tried to make the Background "self-sufficient" so that it is not necessary to read the cited articles. However, if you would like to get any of these documents, let us know which ones you would like and we'll email them to you.

A. Use of Mobile Devices in Online Surveys

In an online survey, it is likely that a sizable percentage of respondents will use, or attempt to use, a mobile device (smartphone or tablet) to complete the questionnaire.

A 2014 AAPOR report on mobile technologies concluded:

> **Recognize If You Are Conducting Online Surveys, You Are Conducting Mobile Surveys:** A non-ignorable and growing percentage of respondents are now accessing online surveys via their mobile browsers (with estimates ranging from 8 - 23% depending on the study), resulting in higher abandonment rates and potentially greater measurement error among these mobile respondents.[38]

With the growth of smartphone ownership and use, the percentages of people completing online surveys on a mobile device has very likely increased since 2014.

[38] AAPOR. (2014). *Mobile technologies for conducting, augmenting and potentially replacing surveys: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research.* Deerfield, IL: The American Association for Public Opinion Research.

In a summary of results from the 2016 General Social Survey (GSS), it was reported by Statistics Canada in *The Daily*[39] that three-quarters of Canadians 15+ own a smartphone, although there is substantial variation by age (we haven't been able to find detailed tables on the Statistics Canada website, so we're stuck with the gaps in age below):

	Smartphone Ownership
Total Canadians 15+	**76%**
15-34	94%
55-64	69%
75 and over	18%

In the 2015 Sage Research report, *Best Practices for Improving Cooperation for Online Surveys*[40], the following are some of the conclusions drawn from a review of the research literature:

- Dropout rates among people using a mobile device to do online surveys are higher than among those using a computer.
- Designing the questionnaire to be mobile-friendly can substantially improve the response rate among smartphone users.
- Another potential contributor to higher dropout rates among people using a smartphone is that it usually takes longer to do a survey on a smartphone than on a computer.
- An issue is whether there are mode effects – e.g. does a mobile-friendly version of a question answered on a smartphone yield the same response distribution as computer-friendly version answered on a computer?

Some survey companies are implementing technologies to accommodate use of mobile devices in online surveys, for example:

- Detection of type of device and browser
- Templates for mobile-friendly question design for various types of questions
- Automatic adaptation of question design to the device, including adapting to screen size

The key point is that adapting questionnaire design for those using a mobile device can potentially improve the survey data in terms of coverage, response rate, reduction of non-response bias, and answer quality.

An online survey can take different approaches to the possibility mobile devices will be used[41]:

1) *Do not adapt the survey to mobile.* This means people attempting to do the survey on a smartphone will not be using a mobile-friendly version of the questionnaire, but rather a version designed for completion on the larger screen of a computer. This will likely produce

[39] Statistics Canada. (2017). "Life in the fast lane: How are Canadians managing?, 2016". *The Daily*, Tuesday, November 14, 2017.
[40] Sage Research Corporation. (2015). *Best Practices for Improving Cooperation for Online Surveys*. Prepared for the Public Opinion Research Directorate, Government of Canada. See pages 53-56.
[41] Callegaro, M. (2010). "Do you know which device your respondent has used to take your online survey?". *Survey Practice*, 3(6).

higher drop-out rates among people using smartphones compared to when a mobile-friendly version of the questionnaire is provided.

2) *Block mobile device users from doing the survey on their device, and encourage them to complete the survey on a computer.* The downside to this approach is that substantial nonresponse could occur if many do not make the effort to switch to using a computer.
3) *Optimize the survey to be correctly displayed on the most common smartphones in use among the survey target group.*
4) *Have the survey be fully compatible to be taken on any device* – which requires a survey platform that can handle adapting the questionnaire to the full range of devices.

The current Standards do not address the possibility and implications of an online survey being completed on a mobile device. The objective for the Panel is to revise the Standards to address these matters.

As we see it, the matters to address in the Standards pertaining to usage of mobile devices to complete online surveys are:

- *1. Proposal Documentation:* Information that must be in the proposal about how usage of mobile devices will be handled in an online survey
- *2. Questionnaire Design:* Requirements and guidelines for a mobile-friendly online survey
- *3. Pre-testing:* Pre-testing the effectiveness of the mobile-friendly versions of the questionnaire
- *7. Data Collection:* Collecting data on type of device
- *14.6. Quality controls:* Look for differences in response distribution if questions are displayed differently on different devices

A1. Proposal documentation relating to use of mobile devices in online surveys

There are proposed revisions to two sections of *Proposal Documentation* relating to use of mobile devices in online surveys:

- 1.2.2 *Sample/Sampling Details*
- 1.2.5 *Questionnaire Design*

Proposed revision to 1.2.2 *Sample/Sampling Details* in Online Standards

The proposal is to add another numbered item to 1.2.2. For now, the number is referred to as "x."

1.2.2. Sample/Sampling Details (in Online Standards)

x) State whether respondents will be able to complete the questionnaire on (a) either a computer or a mobile device (smartphone or tablet), (b) a computer only, or (c) on a mobile device only.
   a) If the intent is to limit completion of the survey to only a computer or to only a mobile device, describe the reasons for this.

   It is likely that without intervention, some people will complete the survey on a computer, and some will use a mobile device. This has implications for questionnaire design because mobile

devices have smaller screens. Restricting device usage can potentially impact coverage of the survey's target population.

Note that item #2 in 1.2.2 (*Describe the sampling procedures*) includes a reference to Section 4, *Sampling Procedures*, where there are requirements to address any issues with representativeness, so it would be redundant to also flag those issues here in item "x" above.

Proposed revision to 1.2.5 *Questionnaire Design* in Online Standards

The proposal is to add another numbered item to 1.2.5. For now, the number is referred to as "x."

1.2.5. Questionnaire Design (in Online Standards)

x) If it is possible the survey will be completed on a mobile device by at least some respondents, then indicate whether or not the intent is to provide a mobile-friendly version of the questionnaire.

A mobile-friendly questionnaire can increase response rates and improve data quality.

a) If the intent is not to have a mobile-friendly version of the questionnaire, describe the reasons for this.

A comment: We expect that people will usually intend to have a mobile-friendly questionnaire, so perhaps the main reason to have this standard is simply to remind people that this is something they need to consider when they get to the questionnaire design stage of the project.

Questions Posted on the Discussion Board

A1.1 Online surveys are likely to be completed on both computers and mobile devices. This is not recognized in the current standards, so the Background proposes various additions to the Standards. Question series A1 to A5 are to get the Panel's input on the proposals, as well as any other suggestions related to use of mobile devices in online surveys.

Firstly, please indicate any suggestions for changes to the proposed addition to *Proposal Documentation,* Section 1.2.2 *Sample/Sampling details*, which is about requiring the proposal to indicate whether the sample will include respondents using either a computer or mobile device.

A1.2 Please indicate any suggestions for changes to the proposed addition to *Proposal Documentation,* Section 1.2.5 *Questionnaire Design*, which is about requiring the proposal to indicate whether there will be a mobile-friendly version of the questionnaire.

A2. Mobile-friendly online surveys and questionnaire design

There are three revisions/additions to Section 2 of the Standards to consider with respect to questionnaire design in online surveys where mobile devices may be used:

- Should there be a standard encouraging use of a common question design/layout across devices
- Section 2.1.1 – questionnaire duration

- Should there be guidelines on features of a mobile-friendly questionnaire, and if so then what guidelines

Should there be a standard encouraging use of a common question design/layout across devices?

The 2015 Sage Research report, *Best Practices for Improving Cooperation for Online Surveys*, reviewed the research literature and concluded:

- Logically, using a different question design/layout on different devices introduces the possibility of a "mode effect" – i.e. a mobile-friendly version of a question answered on a smartphone might yield a different response distribution than a computer-friendly version answered on the larger screen of a computer.
- The research was inconclusive, basically because there hadn't been much research on the topic. It's also the case that it is hard to do conclusive research on this topic, because of the diversity of question types and the diversity of ways in which these can be rendered: research might be conclusive for a particular question type rendered in a particular way on mobile versus computer, but this may not generalize to other question types or to other design approaches.
- That said, there were several studies suggesting that there may not be substantial mode effects[42].

There are two approaches to integrating mobile-friendly question designs into an online survey:

- *Mobile first:* In this approach, the questionnaire is designed from the outset to be mobile-friendly, and the mobile-friendly design is used on both mobile devices and computers. The assumption is that a mobile-friendly design will work just as well on the larger screen of a computer[43]. Note that a mobile-first approach could – and should – use software to adapt the question display to different types/sizes of mobile devices.

- *Responsive design:* Each question design is optimized for mobile and optimized for computer, which means that the question design can be different for mobile (small screen) versus computer (large screen). The idea is that optimizing for each device results in getting the best quality data from each. This approach was recommended in a 2017 research article by Antoun, Couper & Conrad[44]:

  > These results have practical implications for researchers conducting Web surveys. Our finding of near-comparability in the quality of responses between smartphone and PC users,

[42] For example see the following, which include both research reviews and experimental studies:

Toepoel, V., & Lugit, P. (2014). What happens if you offer a mobile option to your web panel? Evidence from a probability-based panel of internet users. *Social Science Computer Review*, *32*(4), pp. 544-560.

Wells, T., Bailey, J., & Link, M. (2014). Comparison of smartphone and online computer survey administration. *Social Science Computer Review, 32*(2), pp. 238-255.

[43] For example:

Saunders, T. (2015). Improving the Survey Experience for Mobile Respondents. *Alert! Magazine*, Third Quarter 2015, Marketing Research Association

Thomas, R. & Barlas, F. (2018). *Best Practices in Mobile-First Design*, GfK Whitepaper

[44] Antoun, C., Couper, M., Conrad, F. (2017). Effects of mobile versus PC web on survey response quality. *Public Opinion Quarterly*, Vol 81, Special Issue, 2017, pp. 280-306

> even for sensitive and burdensome questions, suggest not only that smartphone users should be accommodated in Web surveys but also that the survey instrument should be optimized for small screens. There is a line of thinking, based on *universal mode* design or *unimode* design principles (Dillman 2000), that one should aim to display the exact same question formats across devices in order to minimize mode effects. Our results seem to suggest otherwise since we found no evidence that the redesigned (optimized) question formats in the smartphone version introduced device effects or had an adverse effect on data quality.
>
> The notion of comparability that we would instead argue for is a *best practices* approach … This places the emphasis not on presenting the identical surface-level features, but on using the best practices of each mode or device in order to present the same stimulus to the respondent and minimize error within each mode or device. Carrying out these principles for mobile Web research requires playing to the strengths of smartphones by identifying and using input tools that are user friendly and avoiding any formats that are not (e.g., sliders, drop boxes that turn into pickers). This may also entail using Responsive Web Design (RWD) to adapt a questionnaire to the continuum of different screen sizes rather than binary optimization.

Notably, the same lead author, Antoun, was more cautious in making a recommendation on approach in another article that reviewed research literature from 2007 to 2016 on smartphone optimization:

> Another issue for SO [smartphone optimized] surveys relates to the design and release of different layouts in response to the size of the respondent's screen or web browser. The discussion so far has focused on adapting a questionnaire designed for PCs into a single optimized version for smartphones; but it is important to note that several studies used a responsive design where several versions of the questionnaires were displayed, with the implicit goal of improving response quality within each version (see, e.g., Amin, 2016). This practice has been adopted from web design where a large number of different visual designs and layouts for a single website are deployed (e.g., for small smartphones, large smartphones, small tablets, large tablets, small desktops, and so forth). What is unclear is whether this approach is also effective for surveys where standardization across layouts is a higher priority. A concern is that variations in a particular layout can affect responses (see, e.g., Smyth, Dillman, Christian, & Stern, 2006). Thus, responsive design calls attention to the need to promote comparability across versions, on the one hand, and to minimize error within each version, on the other.
>
> Whether optimization is binary or responsive to a continuum of screen sizes, another issue is whether to design for the biggest or smallest devices first. Almost all of the reviewed studies started from the point of an existing survey that is designed for PCs and then adapted for smartphones. While this approach may prevent major usability problems in the smartphone survey, it does not necessarily produce an optimal design for smartphones (as the word "optimized" implies). Because the smallest screens appear to pose a greater design challenge, the "mobile first" approach may be desirable to the extent that it does not have negative effects on the version of the questionnaire displayed on larger browsers (see, e.g., de Bruijne & Wijnant, 2013a; Tharp, 2015). ***Future research on the effect of responsive design, with and without a mobile-first design philosophy, is necessary before any firm conclusions on these different approaches are drawn.***[45]

The bottom line is the last sentence: there doesn't seem to be a definitive conclusion yet as to which approach is best.

[45] Antoun, C., Katz, J., Argueta, J. and Wang, L. (2017). Design heuristics for effective smartphone questionnaires. *Social Science Computer Review*, Online First

Questions Posted on the Discussion Board

A2.1 Should there be a standard encouraging a particular approach to question design/layout across devices (computers and mobile devices), and if so, what should it be?

Options could be:

- *No standard* – e.g. because not enough is known about what approach is best, so leave it up to the researcher to decide what is best for their project
- *Mobile-first* – which uses a mobile-friendly design for all devices, large and small, and minimizes the possibility of "mode" effects
- *Responsive design* – optimize question design/layout for mobile (small) and computer (large) screens, which can lead to different design/layout across devices, but gives the best data quality for each device
- Don't state either of the above as standards, but rather as guidelines on options to consider

Section 2.1 #1. Questionnaire duration

Because completion times tend to be longer on a mobile device than on a computer, a frequent recommendation is that a mobile-friendly questionnaire should be "short." However, there is no consensus on what this means in terms of number of minutes.

The current Standard for online questionnaire duration is 20 minutes, but an average duration of 15 minutes or less is "strongly encouraged":

2.1. Standards

1) Survey questionnaires must be designed:
   c) to be completed in a maximum duration of 20 minutes, not including pauses or interruptions. Exceptions could include projects with specialized audiences and those with pre-arranged interviews when the respondent is aware the survey will take longer than 20 minutes. **Average questionnaire durations of 15 minutes or less are strongly encouraged in order to minimize respondent burden.**

Questions Posted on the Discussion Board

A2.2 The current standard 2.1.1 c for online questionnaire duration is 20 minutes, but an average duration of 15 minutes or less is "strongly encouraged." Should this be left as is (applies to all online surveys, including mobile-friendly), or should there be a different standard for mobile-friendly surveys?

2.1. Should there be guidelines on features of a mobile-friendly questionnaire, and if so what guidelines?

It is difficult to state "standards" – i.e. requirements – for what constitutes a mobile-friendly question design:

- Asking questions that effectively address survey objectives is the top priority, and this may constrain design choices.
- The designs supported by survey software varies across vendors.
- Similarly, software to dynamically adapt question design to device type varies across vendors.
- There is researcher experience and judgment that goes into design choices.

One option in the Standards is not to give any standards or guidelines on mobile-friendly design. The idea is that in *Proposal Documentation* there will be a statement of intent to have a mobile-friendly survey. Beyond that, it will be up to the researcher to decide how to implement that for their survey. And, under this view, guidelines do not belong in a Standards document.

The other option is to give guidelines – i.e. recommended, but not required, practices or principles. This is certainly possible to do based on the research literature. The idea is that it would be helpful to GC POR researchers to have this guidance stated in the Standards document.

For reference, two approaches for guidance are shown here.

1. The Sage Research report, *Best Practices for Improving Cooperation for Online Surveys*, summarized the following guidelines based on a review of the literature (detailed citations are in the report):

✓ *Write short question text and limit the number of response options.*	The goal is to avoid/minimize scrolling to see the entire question and all of the response options.
✓ *Use large fonts and allocate sufficient space for touch selection of response options.*	This will make it easier to read and respond when using a touch interface, and reduce response errors.
✓ *Avoid requiring zooming to be able to see the question and response options.*	Requiring zooming increases respondent burden, and the time needed to complete the survey.
✓ *Avoid visual clutter – i.e. visual elements that do not directly support efficient presentation of the question and response options.*	These are distracting, and can increase page load times.
✓ *Minimize images, and other high bandwidth requirements.*	Note also that Flash content – which can be used to present video – is not supported on devices using the iOS operating system (iPhones, iPads).
✓ *Use page breaks carefully.*	On the one hand, spreading questions across different pages reduces page load time; on the other hand, too many pages can make for a slow experience waiting for pages to load. Orr (2012) suggests a maximum of three questions per page as a rule of thumb; AAPOR (2014) suggests a maximum of two questions per page. However, Mavletova and Couper (2014) found directionally higher break-offs when there was only one question per page compared to multiple questions

	per page, suggesting it is better to have more than one question per page.
✓ *Consider limiting the requirement for typing/text entry.*	AAPOR (2014) summarizes evidence indicating this has been an issue in the past, but hypothesize that this potential limitation may be going away, as people become more used to texting on their phone.
✓ *Use small grids that require little or no horizontal scrolling, or avoid grids altogether.*	
✓ *Avoid horizontal scrolling; vertical scrolling is more acceptable.*	For example, both Peytchev and Hill (2010), and Stapleton (2013) found that those using a mobile device were more likely to pick the immediately visible response options rather than scroll horizontally to see more options.
✓ *Avoid interactive elements such as sliding scales or drag-and-drop as these are usually too difficult to use on a mobile device.*	

To this list we would add:

✓ *Use survey software that detects the device being used and can automatically adapt question design/layout to the device.*	

2. Another option for guidelines is the five "design heuristics" proposed by Antoun et al (2017)[46] based on their review of the literature:

✓ *Readability*	Text should be large enough to promote easy reading.
✓ *Ease of selection*	Touch targets should be large enough to tap accurately.
✓ *Visibility across the page*	All content should fit the width of the screen so that it is visible without horizontal scrolling.
✓ *Simplicity of design features*	Design features should be simple both for researchers to deploy and for respondents to use. Deploying complicated features (sliders, pickers, questions with embedded video) can increase the occurrence of technical problems. Even without glitches, complicated features can increase missing data or harm data quality if some

[46] Antoun, C., Katz, J., Argueta, J. and Wang, L. (2017). Design heuristics for effective smartphone questionnaires. *Social Science Computer Review*, Online First

	respondents do not understand or are not able to learn in real time how to use the features. [Note: a "picker" is a rendering for mobile of a drop-down list]
✓ *Predictability across devices* [Comment: We think "Displays appropriately across devices" would be a better label.]	Questionnaires should function in a predictable way across different devices. The concern is that questions will be rendered differently across different types of platforms (e.g., pickers vs. spinners). Another concern involves different screen sizes. Touch targets and font sizes may meet design standards on big devices (e.g., 6-in. displays) but not on smaller devices (e.g., 4-in. displays).

Questions Posted on the Discussion Board

A2.3 In Section 2 of the Standards, *Questionnaire Design*, should there be guidelines (recommendations, but not requirements) in the Standards on what makes a questionnaire mobile-friendly, or is that not necessary/appropriate in the Standards?

In the event guidelines are included in the Standards, what do you suggest these be? Possibilities were outlined in the Background. It would also be helpful if you could take a crack at writing them as you would like them to appear in the Standards.

A3. Proposed revisions to *3. Pre-testing* in the Online Standards

There are two types of pre-testing that should be done for an online survey questionnaire where some respondents may use a mobile device:

1) Internal pre-testing by the researchers to see whether the questions display appropriately on a sample of different devices that respondents might use
2) External pre-testing with respondents using different devices

The current Standards do not specifically refer to requirements for online surveys where mobile devices may be used.

In Section 3 of the Online Standards, the two most relevant current standards are:

3. PRE-TESTING

3.1. Standards

1) Pre-testing of all components of a new or revised survey questionnaire that may influence data quality and respondent behaviour is required. This includes the online appearance and functionality of the questionnaire.

5) A minimum of 10 pre-test surveys are to be completed in each language in which the final survey will be fielded. An exception could be projects with small survey populations, such as a client-satisfaction survey of a small client base. In such cases the researcher must, in consultation with the client, take steps to ensure that the smaller number of pre-tests are

sufficient to guarantee questionnaire quality. For example, a cognitive pre-test may be warranted.

Should standards specific to completion of an online survey on mobile devices be added, or not?

An argument for not adding anything is that Standard 3.1.1 is sufficiently broad in scope that a requirement for pre-testing of the mobile-friendly aspects of the survey is implied even if not explicitly stated.

An argument in favour of adding standards is to make sure the pre-testing is done appropriately – with the added standards stating the minimum requirements.

What, if any, revision should be made to Standard 3.1.1?

For example, a revision could be:

1) Pre-testing of all components of a new or revised survey questionnaire that may influence data quality and respondent behaviour is required. This includes the online appearance and functionality of the questionnaire.
    a) For mobile-friendly surveys, researchers must also do internal pre-testing on a sample of the types and sizes of devices that respondents might use. On smartphones, the internal pre-testing must look at how questions display in both portrait and landscape mode.

What, if any, revision should be made to Standard 3.1.5?

For example, a revision could be:

5) A minimum of 10 pre-test surveys are to be completed in each language in which the final survey will be fielded. An exception could be projects with small survey populations, such as a client-satisfaction survey of a small client base. In such cases the researcher must, in consultation with the client, take steps to ensure that the smaller number of pre-tests are sufficient to guarantee questionnaire quality. For example, a cognitive pre-test may be warranted.
    a) For online surveys the pre-test must include respondents using at least some of the different device types and sizes that might be used to complete the survey. For example, in a survey that can be completed on either a mobile device or a computer, the sample must include a specified minimum number of both mobile device users and computer users.

Questions Posted on the Discussion Board

A3.1 Question A3.1 to A3.3 pertain to the possibility of adding standards to Section 3, *Pre-testing*, to address pre-testing of an online survey questionnaire on both mobile devices and computers.

Should standards specific to completion of an online survey on mobile devices be added, or not?

A3.2 The Background gives a possible addition to 3.1.1 that requires internal pre-testing of an online survey questionnaire on *a sample of the types and sizes of devices that respondents might use*. Should this be added to 3.1.1? Any suggestions for revisions?

A3.3 The Background gives a possible addition to 3.1.5, on number of pre-test completions, to address the number of pre-test completions on mobile devices and computers. Should a standard of this sort be added to 3.1.5? Any suggestions for revisions?

A4. Possible revisions to *7. Data Collection* and *14.6 Quality Controls* in the Online Standards related to the possibility of mode effects by device type/screen size

In a survey that allows completion on both mobile devices and computers, and particularly one using a "responsive design" approach that can result in different question designs/layouts for different size screens, there is the potential for a "mode" effect. That is, the different designs/layouts for a given question could cause different response distributions. If this possibility is to be explored, then data on device type needs to be collected during the survey.

Also, there is the possibility of a device-type effect as a result of differences in the characteristics of people who use a mobile device to complete a survey versus people who use a computer. For example, a research vendor noted the following in an article advocating for "device agnostic" sampling for online surveys:

> Specific pockets of the population gravitate toward mobile and we expect to see the level of systematic non-coverage bias to grow in non-mobile research designs. Some of our testing demonstrates between 20-25 percent of millennials (those born from 1981-2000) prefer to access surveys via mobile, which means our non-mobile surveys are missing the views of a substantial portion of this audience. Considering that this group is likely more tech savvy and connected with peers, we expect over time to see biased and inaccurate results when they are excluded from our sampling frame.[47]

This illustrates the importance of a mobile-friendly survey for ensuring good coverage of the population. However, it also means that an exploration of possible effects of question design/layout would need to be done carefully to avoid confounding question design/layout with covariates such as age.

As noted earlier, the conclusions from the research literature are:

- There tend not to be significant "mode" effects.
- But, the research literature to date is inconclusive because of the limited number of studies and the limited question design variations studied.

Questions Posted on the Discussion Board

A4.1 Should the Standards have a requirement (or guideline) for researchers to look at whether different question designs/layouts in smartphones versus computers affected response distributions in the survey data?

A4.2 If analysis of device type/screen size effects is required or encouraged in the Standards, how should Section 7.7, *Monitoring of Online Survey Fieldwork*, be revised? What types of

[47] Simpson, S. (2014). Okay fine, I'll go device agnostic. *Survey Magazine*, June 2014.

information on devices should be gathered? Should the information be collected for all online surveys, or only those with certain characteristics (e.g. different question design/layout by device type)?

For example, a possible addition to 7.7 could be the following guideline (or should it be a standard – i.e. a requirement?):

7.7. Monitoring of Online Survey Fieldwork

3 If possible, the survey system should collect data to allow classification of a device as a smartphone, tablet or computer. If this is not possible, there should be a question in the survey to gather this information.

A4.3 If analysis of device type/screen size effects is required or encouraged in the Standards, how should Section 14.6, *Quality Controls*, be revised?

Section 14.6 already has an analogous requirement for multi-mode surveys:

14.6. Quality Controls

3) For multi-mode surveys, detailed description of any data quality issues arising from combining data collected via different modes/instruments.

A possibility could be:

4) For online surveys where there are substantial differences in question design/layout by device type, detailed description of any data quality issues arising from combining data collected from the different question designs/layouts.

A5. Any other revisions to Standards associated with mobile-friendly online surveys?

Questions Posted on the Discussion Board

A5.1 To summarize, possible revisions to the Standards pertaining to use of mobile devices in online surveys have been discussed for:

- *1. Proposal Documentation:* Information that must be in the proposal about usage of mobile devices will be handled
- *2. Questionnaire Design:* Requirements and guidelines for a mobile-friendly survey
- *3. Pre-testing:* Pre-testing the effectiveness of the mobile-friendly versions of questions
- *7. Data Collection:* Collecting data on type of device
- *14.6. Quality controls:* Look for differences in response distribution if questions are displayed differently on different devices

Are there any other areas where standards or guidelines should be added or revised with respect to usage of mobile devices in online surveys? Note that Section B of the discussion board will deal with covering respondent costs for using a mobile device.

B. Use of Mobile Devices – Covering Respondent Costs

Users of mobile devices may incur costs to participate in a research survey.

The current standards do not have any requirements as to how such costs should be handled.

The following is a standard/guideline in the *ESOMAR/GRBN Guideline on Mobile Research:*

> **3.1.3 Costs**
>
> Unlike most other research methods, data subjects may incur costs as a consequence of participating in mobile research that may include charges for data downloads, online access, text messaging, data plan overages, roaming charges, voicemail message retrieval and standard telephone charges. Researchers should design their research so that data subjects incur no costs without express approval. If this is not possible, researchers must be prepared to offer compensation. Such compensation may be cash, mobile money, airtime or other forms of value.

Note the ESOMAR/GRBN explain their use of "must" and "should" as follows:

> Throughout this document the word "must" is used to identify mandatory requirements. We use the word "must" when describing a principle or practice that researchers are obliged to follow. The word "should" is used when describing implementation. This usage is meant to recognise that researchers may choose to implement a principle or practice in different ways depending on the design of their research.

The objective for the Panel is to determine whether something like the ESOMAR/GRBN 3.1.3 above should be incorporated into the Online and Telephone Standards.

The following is slight re-wording of the ESOMAR/GRBN – we suggest you use this as a starting point for making suggestions:

> **Proposal:**
>
> Respondents using a mobile device may incur costs as a consequence of participating in mobile research that may include charges for data downloads, online access, text messaging, data plan overages, roaming charges, voicemail message retrieval and standard telephone charges. Researchers should design their research so that data subjects incur no costs without their express approval. If this is not possible, researchers must be prepared to offer compensation. Such compensation may be cash, mobile money, airtime or other forms of value.

Questions Posted on the Discussion Board

B1. Should the Online and Telephone Standards have a section about covering respondent costs of using a mobile device? If so, what do you think of the proposed text in the Background based on a ESOMAR/GRBN standard? Please revise as you see fit.

C. Proportion of Cell versus Landline Phones in Probability Telephone Surveys

An important issue in sampling for telephone surveys is the inclusion of cell phone users and landline users. This can affect coverage of the survey population, the sampling frame(s) used for the survey, and possibly weighting. The objective for the Panel is to identify any revisions to standards pertaining to this issue. Of particular interest is what, if anything, to say about the proportion of cell phone versus landline interviews in a probability survey of the Canadian population.

The following are some statistics taken from the CRTC's *Communications Monitoring Report 2017*:

- The data are from Statistics Canada's 2015 Survey of Household Spending, so these are outdated. Very likely, the percentage of cell phone only has increased since then, and landline-only and dual ownership has decreased.
- Cell phone only households in Canada = 27.5%, landline-only = 13.2%, and dual cell-phone/landline = 58.6%
- The types of phones in households varies demographically:
  - -- The percentage of cell phone only households varies by household income, being highest in the lower-income households: lowest income quintile = 35.0%; highest income quintile = 18.8%
  - -- The percentage of cell phone only households varies by province, from 14.4% in New Brunswick to 34.7% in Alberta.

  The CRTC report does not report phone type by age, but back in 2010, the AAPOR Cell Phone Task Force concluded: *"young adults in the U.S. aged 18 to 34 years, can no longer be reached successfully via the landline frame."*[48]

These data indicate that a telephone probability sample of the general Canadian adult population must include a cell phone sample.

For this section, please review *3. Dual-Frame Probability Telephone Surveys*, in the literature review commissioned by PORD (pp. 12-14, and recommendation on pp. 31-32 for *Sample frame proportions for dual-frame surveys*). Some points made in the literature review:

- No industry associations have set standards for numeric proportions of landline and cell phone numbers.

  As an example of what a respected survey organizations does currently in the U.S., the literature review cites Pew Research, which now targets to complete 75% of interviews by cell phone and 25% by landline[49]. The literature review notes, though, that *"cellphone-only*

[48] AAPOR Cell Phone Task Force. (2010). *New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. with Respondents Reached via Cell Phone Numbers.* This is a very good review of basic issues that need to be considered when including a cell phone sample in a telephone survey.

[49] Pew Research, *Our Survey Methodology in Detail*, http://www.pewresearch.org/methodology/u-s-survey-research/our-survey-methodology-in-detail/

households have not reached the same level of penetration in Canada as compared to the United States."

- The optimal proportion of cell phones versus landline phones will vary depending on the population being surveyed: *"The optimal allocation for a dual-frame telephone survey depends on the proportion of the population that only has a cellphone and the target audience. … For example, it will be more cost effective to conduct a survey of young adults almost exclusively using cellphone sample…"*
- It is premature to use only a cell phone sample frame to get a good probability sample in Canada. Rather, it is necessary to use a dual frame design – i.e. both a cell phone sample frame and a landline sample frame.
- With regard to weighting a dual-frame sample: *"At the time of writing, there is no consensus on the best approach to weighting dual-frame survey samples."*

  The literature review suggests that in Canada, often the following approach is used:

  > In Canada, when conducting a telephone survey of the general public, it is common to apply one set of demographic weights to survey data thereby treating cellphone and landline interviews the same. A dual-sample frame is used to minimize the potential for coverage error resulting from the incomplete landline sample frame … From this perspective, survey samples resulting from dual-frame sampling only require weighting adjustments to bring the demographic distributions in line with those of the population. (p.14)

- The literature review recommends: (a) don't set numeric standards for the proportion of cell phone and landline numbers, and (b) have a "guideline" (not a standard/requirement) that survey data reflect the population of cell phone only households:

  > Sample frame proportions for dual-frame surveys: The proportion of landline and cellphone numbers included in dual-frame samples varies as researchers try to find the optimal balance and as the incidence of exclusive cellphone use increases. For this reason, it is recommended that PORD not put in place prescriptive standards for dual-frame sample proportions. There will not be one ratio of landline-to-cell sample that will apply to all public opinion research studies. Other factors, such as budget, target audience, and field period, will influence the proportions.
  >
  > A useful guideline would be one that focusses on the outcome of the data collection. That is, the sample ratios chosen should ensure that the proportion of cellphone-only households in the final survey sample reflects that of the population of cellphone-only households at the time of the study.

The 2010 AAPOR Cell Phone Task Force noted a variety of other issues involving surveying cell phone users, and typically noted that there is no definitive "best" resolution. For example:

- How to deal with the fact that a cell phone sample frame overlaps with a landline sample frame (i.e. dual users are in both)
- How to handle dual users who are "cell phone mostly" or "landline mostly"
- Selection of respondent in a household when dialing a cell phone (e.g. Pew Research selects the person answering the cell phone, but for landline has a process for selecting someone in the household who may not be the person answering the phone)
- Selection of respondent when there is cell phone sharing within a household

- How to treat a business cell phone that also may be used sometimes for personal matters

The following are the current standards in the Telephone Standards that refer to cell phones:

1. PROPOSAL DOCUMENTATION

1.2. Technical Specifications of the Research

1.2.3. Response Rate/Participation Rate and Error Rate

1) State the expected response rate/participation rate for the sample(s). State expected response rates/participation rates for each mode when more than one mode is proposed (i.e. separately for online and telephone components). Similarly, if a component of the research will recruit respondents via their **cell phones**, the expected response rate/participation rate for that component must be stated separately as well. The expected response rate/participation rate is an estimation based on various factors such as previous response rates/participation rates, trends etc. The proposal must include a brief discussion of the factors that might cause the actual response/participation rate to fall short of the stated target.

1.2.4. Description of Data Collection

7) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method. When it is proposed to recruit at least some respondents via their **cell phones**, provide a rationale for using this approach in addition to or in place of landline phones.

4. SAMPLING PROCEDURES

4.2. Probability Sampling

3) A full description of the sample design and selection procedures must be stated including:

   c) at each sampling stage, the method of achieving a probability sample (e.g., random selection) shall be explained, and any subsets of the universe that have been excluded or underrepresented shall be stated (e.g., **cell phone only** households), although whenever possible, an estimate of the percentage of the universe that has been excluded or underrepresented must be provided;

Note that Standard 1.2.4 is repeated in Standard 15.5 *Data Collection* (telephone) as part of Section 15, *Mandatory Survey Report Requirements*.

Questions Posted on the Discussion Board

C1. **Section 1 Proposal Documentation, Standard 1.2.3 #1 on response rate/participation rate:** This standard states that if a telephone survey sample includes cell phone users, the expected response/participation rate should be stated for that component. It does not explicitly state the same requirement for landline users, although that may be implied.

Any suggested changes to the current Standard 1.2.3 #1 with respect to stating expected response rate , or is it fine as is?

C2. **Section 1 Proposal Documentation, Description of Data Collection, Standard 1.2.4 #7:** This standard states that a rationale must be given when the sample includes interviews on cell phones. The language arguably overly downplays the importance of including cell phone users

in the sample. The following is a proposed revision based on the literature review commissioned by PORD (leaving aside the "multi-mode" part of #7):

> **1.2.4 #7:** The telephone survey should (must?) include interviewing by both cell phone and landline. The sample ratios of cell phone to landline must be stated, and should (must?) ensure that the proportion of cell phone only households in the final survey sample reflects that of the population of cell phone only households at the time of the study.
>
> If interviewing is to be done only by landline or only by cell phone, provide the rationale.

- Any suggested changes?
- Can/should anything more specific be stated about the numeric proportions of cell phone and landline sample in a probability telephone survey in Canada?
- Should this be worded as guidelines ("should") or standards ("must")?

C3. **Section 1 Proposal Documentation & Section 4 Sampling Procedures:** These sections of the Telephone Standards do not have any requirements with respect to how cell phone and landline sampling will be done, nor to any post-survey weighting specifically based on cell phone/landline status.

There may not be a need for this level of specificity, because the general requirements are sufficient to trigger disclosure of the information (i.e. the researchers will translate the general requirements into the information that needs to be stated about cell phone/landline for the particular study in question). Alternatively, perhaps the Standards should address the cell phone/landline topic in a more specific way.

Are there any additional design matters or information disclosure requirements pertinent to the cell phone/landline topic that you think should be added to either Section 1 Proposal Documentation or Section 4 Sampling Procedures? If so, please indicate whether these should be standards (requirements) or guidelines (suggestions).

C4. **Section 4 Sampling Procedures, 4.2 Probability Sampling, Standard 4.2.3c:** This gives an example focusing only on cell phone only households. Should the example be expanded to include landline only households? That is:

> **4. SAMPLING PROCEDURES**
>
> **4.2. Probability Sampling**
>
> 3) A full description of the sample design and selection procedures must be stated including:
>
> c) at each sampling stage, the method of achieving a probability sample (e.g., random selection) shall be explained, and any subsets of the universe that have been excluded or underrepresented shall be stated (e.g., cell phone only **or landline only** households), although whenever possible, an estimate of the percentage of the universe that has been excluded or underrepresented must be provided;

D. Telephone Survey Call-Back Requirements

The Telephone Standards for call-backs are in Section 7, *Data Collection*:

> **7.2. Call-Backs**
>
> 1) There will be a minimum of eight call-backs made before retiring a telephone number and substituting it with another number. The call backs must be made at varying days and times over a minimum seven-day period. An exception could be made when the field period is shorter as a result of the need to assess recall of particular events or activities.
> 2) Every effort must be made to ensure that the respondent is called back if an appointment has been arranged and that the date and time of that appointment are respected.
> 3) No attempt will be made to call back refusals.

The Panel is being asked for input on:

- Is *"a minimum of eight call-backs"* the appropriate number?
- Should there be a different call-back standard for IVR surveys?

D1. Is *"a minimum of eight call-backs"* the appropriate number?

PORD posed this question to the MRIA, as concern had been expressed about whether eight call-backs is too many, and might be perceived as harassment.

The MRIA *Polling Standards for the Canadian Marketplace*[50] state eight call-backs as a maximum, as compared to Section 7.2 *Call-Backs* which states it as a minimum. The standards also state a definition of call-backs (no definition is given in Section 7.2 of the Telephone Standards).

> Make no more than eight (8) calls to the same telephone number. This number includes:
>
> - callbacks made to establish initial contact with the potential respondent (for instance, when there was no answer or the line was busy on previous calls); and
> - callbacks made after contact has been established (for instance, when the potential respondent asked for a callback at a more convenient time, or when the selected respondent was not home)

MRIA posed the following question to some members: *The MRIA had adopted the policy of a maximum of 8 calls to each potential respondent. See Appendix "L" 8.4.2 . If the number of calls allowed are reduced would this affect your research studies? What is your firm's frequency of callbacks - Over what period do you make the (eight) 8 calls?* The response relayed to PORD was:

> Members are very sensitive to respondent fatigue and aim to regulate the frequency of calls to the same respondents.

[50] MRIA Code of Conduct for Market and Social Research, Appendix L, *Polling Standards for the Canadian Marketplace*

The additional calls to respondents are usually after exhausting the list of potential respondents and calls are made to those whom the researcher was unable to contact in the initial call.

It is extremely important to recognize the difference between dispositions for call attempts. Not every call attempt should be considered a call-back. For example, 8 call backs, all yielding a busy signal, is very different than 8 call backs all resulting in an answer and call back request. At the same time, the number of call backs must be large enough to provide a reasonable expectation of equal probability of selection for all primary sample units. We have a maximum of 7 call backs, but do consider some call dispositions to be partial callbacks (e.g., a busy signal counts as 1/3 of a call-back). So theoretically we could call a phone number up to 21 times (21 busy signals). Usually phone numbers are resolved after 6 - 10 attempts.

Refusal conversion dialing must also be considered and whether these conversion attempts are considered within the call back limit.

PORD also would like the Panel to comment on whether the call-back requirements should be the same for respondents using a cell phone.

Questions Posted on the Discussion Board

D1.1 Section 7.2 of *Data Collection* on *Call-Backs* specifies a minimum of eight call-backs. As indicated in the Background, the MRIA specifies eight as the maximum, and some MRIA member feedback implies some variability in how a call-back is defined.

Do you suggest any changes to Section 7.2 *Call-Backs* in terms of the number of call-backs specified? Also, should the standard apply in the same way to respondents reached on a cell phone, or should they be treated differently?

D1.2 Section 7.2 *Call-Backs* does not state a definition of call-backs. Should it, or is this not necessary? If so, what should it be? For reference, the MRIA *Polling Standards* said call-backs include:

- callbacks made to establish initial contact with the potential respondent (for instance, when there was no answer or the line was busy on previous calls); and
- callbacks made after contact has been established (for instance, when the potential respondent asked for a callback at a more convenient time, or when the selected respondent was not home)

D2. Should there be a different call-back standard for Interactive Voice Response (IVR) surveys?

The call-back requirements in Section 7.2 *Call-Backs* do not make any distinction between interviewer-administered surveys and IVR surveys.

7.2. Call-Backs

1) There will be a minimum of eight call-backs made before retiring a telephone number and substituting it with another number. The call backs must be made at varying days and times over a minimum seven-day period. An exception could be made when the field period is shorter as a result of the need to assess recall of particular events or activities.

In the literature review commission by PORD, under *Drawbacks of using IVR*, it states:

> In addition to limits on the length and complexity of the survey questionnaire, the quality of the sample can be questionable. This, however, is not unique to IVR. All survey research requires complete sampling frames (little to no coverage error), sound sampling strategies (simple random sampling or stratified random sampling), and appropriate sample control (an adequate number of call-backs that vary by time of day/day of the week to maximize the response rate). Since speed is one of the key advantages of using IVR, sample control measures are not as stringent (there is no time to call back a number in the sample multiple times). The survey sample will include whoever could be reached on the night of the data collection, which introduces the possibility of non-response bias. While survey weights will be applied to the survey sample post-data collection to ensure it reflects the demographic profile of target population, this will not address attitudinal differences that might exist between survey respondents and non-respondents. (p. 18)

The literature review implies that it may often be the case that few, if any, call-backs are made in an IVR survey. It connects this to using IVR for its speed advantage, and note that Standard 7.2 #1 includes an exemption from the eight call-back requirement when speed of fieldwork is important: *"An exception could be made when the field period is shorter as a result of the need to assess recall of particular events or activities."*

Questions Posted on the Discussion Board

D2.1 Should any there be any changes to Section 7.2 *Call-backs* specific to IVR surveys?

E. IVR Surveys

In an Interactive Voice Response (IVR) telephone survey, a computer is programmed with a questionnaire, calls are made automatically, and the recorded questions are read by the computer. There are no live interviewers.

The objectives for the Panel are:

- Suggest any needed revisions to Section 5.3 of the Telephone Standards, *Use of Interactive Voice Response*
- Comment whether the standard for survey duration should be different for IVR surveys

For this topic, please review the section *Interactive Voice Response Surveys* in the literature review commissioned by PORD (pp. 17-19). To summarize, the literature review identifies the following:

Advantages of IVR

- More time and cost effective compared to live interviewer surveys
- Because of its lower cost, IVR can be useful for surveying low incidence populations
- Less subject to social desirability effects
- Have done well in some election forecasting surveys in Canada

Disadvantages of IVR

- Works less well for lengthy surveys due to hang-ups, so it is better suited to short surveys
- Questions about sample quality: the literature review specifically mentions an IVR design in which there are no call-backs, and the *"survey sample will include whoever could be reached on the night of the survey, which introduces the possibility of non-response bias."* The review notes that there is no guarantee that weighting will eliminate non-response bias.
- Negative perceptions of the methodology, which may be particularly important for Government of Canada public opinion research

E1. Section 5.3 *Use of Interactive Voice Response*

Section 5.3 *Use of Interactive Voice Response* in the Telephone Standards discourages, but does not forbid, use of IVR surveys for POR. It also suggests circumstances when IVR may be an appropriate methodology. The standard states that IVR surveys have the same requirements as interviewer surveys for the survey introduction, respondent opt-out, times when calls can be made, and delay in acknowledging an answered call. The standard does not explicitly state that call-back requirements apply to IVR surveys (see D3 for a question about this).

5. RETAINING PUBLIC CONFIDENCE

5.3. Use of Interactive Voice Response

1) Characteristics of Interactive Voice Response (IVR) surveys, including the impersonal style conveyed by automation, put that method in conflict with the manner in which the Government of Canada wishes to engage Canadians. IVR can therefore be used only when a convincing case is made that the specific information to be collected is essential for making important decisions and cannot be obtained through other means. For example, IVR may be judged acceptable when the opinions of a hard-to-reach (low incidence) group are critical to the issue at hand and the very high call volume made economical by IVR is likely to markedly increase participation from members of that group in the survey. (IVR may also be used whenever respondents have agreed beforehand to this method.)
2) When IVR is used, the same information required for interviews conducted by live interviewers (sponsor, researcher, participation is voluntary, assurance of confidentiality, etc.) must be included in the IVR survey introduction. Respondents must also be provided early in the introduction with an easy method to opt out of the survey (e.g., by pressing a specific key) so that the call is terminated gracefully and no more calls are made to that number. The same requirements for the time-of-day of calls and delay in acknowledging an answered call (see section 5.2.) apply to the use of IVR.

Note that the relevant parts of Section 5.2 referred to above are as follows. **Note also that the next Discussion Board will take another look at #2 (hours of calling) for telephone surveys generally, so our interest here is only in issues specific to IVR surveys.**

5.2. Avoidance of Harassment

2) Using the respondent's local time, telephone interviewing can only be conducted from 9:00 to 21:00 Monday to Friday, 10:00 to 21:00 Saturday, and 12:00 to 21:00

Sunday and statutory holidays except when a pre-arranged interview is requested by the respondent outside these hours.

3) There must be no or minimal (one second) pause before the interviewer acknowledges that a potential respondent has answered the telephone.

Questions Posted on the Discussion Board

E1.1 Do you have suggestions for revisions to Section 5.3 *Use of Interactive Voice Response*? Considerations could include:

- Overall stance towards use of IVR versus interviewer surveys
- Examples of when use of IVR would – or would not – be appropriate
- In 5.3 #2 the listing of interviewer survey standards that apply to IVR surveys (we're considering call-back requirements separately, in the next section)
- Anything else you think should be in Section 5.3 either as a standard or as a guideline

E2. IVR survey introduction

The standard for survey duration for telephone states surveys must be completed in 20 minutes, and strongly encourages a duration of 15 minutes or less.

2. QUESTIONNAIRE DESIGN

2.1. Standards

1) Survey questionnaires must be designed:

c) to be completed in a maximum duration of 20 minutes. Exceptions could include projects with specialized audiences and those with pre-arranged interviews when the respondent is aware the survey will take longer than 20 minutes. **Average questionnaire durations of 15 minutes or less are strongly encouraged in order to minimize respondent burden.**

The literature review commissioned by PORD suggests that IVR surveys are better suited to shorter surveys:

> One of the main drawbacks of using the IVR methodology is the need for a short, simply constructed questionnaire. This is one of the reasons this methodology is well-suited to election polling and measuring voter intention. One or two clear and concise questions can be asked when using IVR and these questions will have simple (and few) response options from which to select. For example, "press 1 if you know who you intend to vote for on election day and 2 if you do not".
>
> For a research project that intends to explore several topics with respondents and/or topics in a more in-depth manner (e.g., to uncover reasons for voter intentions or to gain insights on salient election issues with voters), an IVR survey would not be the appropriate methodology. (p.18)

PORD has found that typically IVR surveys are less than 20 minutes. PORD also noted that they have seen IVR being used more in the past year for recruiting. In this context, PORD has raised a question about whether the required elements for telephone survey introductions should be revised or

shortened for IVR surveys. Note that exempting IVR surveys from any of these elements would be significant, and would need strong justification. The current standard is:

2. QUESTIONNAIRE DESIGN

2.1. Standards

2) The following are required elements of all Government of Canada telephone survey questionnaire introductions:

a) follow the *Official Languages Act* and Policies; in particular, introductions must include an active offer of both official languages (English and French); the language that is used first will depend on the province in which the respondent resides (e.g., in Quebec "Bonjour/Hello"; in the Rest of Canada "Hello/Bonjour");

b) identify the Government of Canada or the department/agency sponsoring the survey;

c) inform respondents of the general subject and purpose of the study. The subject and purpose of the study may be expressed in very general terms so long as these terms cannot be construed as an attempt to misinform respondents. When the researcher makes a convincing case that the subject and purpose of the survey is such that stating it at the outset will affect respondents' willingness to participate, stating them may be deferred to the conclusion of the study;

d) inform respondents of the expected length of the interview;

e) identify the researcher and interviewer (a pseudonym may be used for the interviewer as long as the individual remains identifiable by management internally for quality control purposes);

f) inform respondents that their participation in the study is voluntary and completely confidential*;*

g) inform respondents that their responses remain anonymous. In the exceptional cases where the research objectives require that respondent identity be revealed, the informed consent of the respondent must be obtained;

h) inform respondents that the survey is registered with the Research Registration System maintained by the MRIA and provide information sufficient for respondents to access the System's Research Verification Service (project registration number and service's website URL, e-mail or phone number), if requested.

Examples of exceptions: The sponsor would not be mentioned until the end of the survey in particular cases (e.g., advertising post-tests); other languages would be mentioned when the survey is available in non-official languages; and information about the availability of reports could be provided to specialized respondents when the incentive to participate outweighs the additional length of the introduction.

Note: If requested by the respondent, interviewers need to be prepared to repeat the instructions. Interviewers must have information readily available to inform respondents about how respondents were chosen, how privacy is protected, and where the survey results can be obtained.

The current standard for IVR surveys specifies that all of these items must be included in an IVR survey:

5.3. Use of Interactive Voice Response

2) When IVR is used, the same information required for interviews conducted by live interviewers (sponsor, researcher, participation is voluntary, assurance of confidentiality, etc.) must be included in the IVR survey introduction. …

Questions Posted on the Discussion Board

E2.1 PORD has raised a question about whether the required elements for telephone survey introductions should be revised or shortened for IVR surveys (see Background for listing). Note that exempting IVR surveys from any of these elements would be significant, and would need strong justification. Should any of the required elements be omitted or revised for IVR surveys?

E2.2 PORD has asked that the Panel comment on the following type of introduction for an IVR survey. This approach moves some of the required information disclosures to the "end of the survey", where the information can then be accessed by calling a number or going to a website:

> ***Introduction for an IVR survey:*** *Hello/bonjour, this is (survey company) calling on behalf of the Government of Canada/government department. Pour continuez en francais appuyez sur le (number). We are conducting a (number) minute research survey on SUBJECT. Your participation is voluntary and completely confidential and anonymous. The name, telephone and web address of whom to contact for additional information about this research project are available at the end of the survey. To continue press (number). To end this call now hang up or press (number).*
>
> ***(At the end of the call)*** *If you would like further information on this study please contact (name and phone number) or go to (website). To repeat this information press (number).*
>
> *Further information could include:*
> - *that the survey is registered with the Research Registration System maintained by the MRIA*
> - *how to access the System's Research Verification Service (project registration number and service's website URL, e-mail or phone number)*
> - *how respondents were chosen, how privacy is protected, and where the survey results can be obtained.*

What do you think of this type of introduction for an IVR survey, which for IVR surveys (but not interviewer surveys) moves some information disclosures to the "end of the survey" in a form that requires the person to take an additional action to get that information? Any suggested revisions to the above text?

F. Multi-Mode Surveys

PORD would like to clarify and strengthen the standards for multi-mode surveys.

For reference, the current Standards referring to multi-mode surveys are:

1. PROPOSAL DOCUMENTATION

1.2.4. Description of Data Collection

7) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method.

4. SAMPLING PROCEDURES

4.5. Multi-Mode Surveys

Multi-mode surveys are ones where different methods of questionnaire administration are used. They will often involve a combination of online and telephone methods, although there are other possibilities (e.g., in-person, mail, fax).

When a survey is conducted using multiple modes of questionnaire administration:

1) The reasons for using a multi-mode rather than a single-mode method must be stated, both in the research proposal and the survey report.
2) When the plan is to combine data collected via different modes in the data analyses, then steps must be taken to ensure as much comparability as possible across the different survey modes in terms of question wording and presentation of response options.
3) Steps must be taken to ensure avoidance of duplicate respondents in different modes. The steps taken, and the results, must be documented.

14 (online)/15 (telephone). MANDATORY SURVEY REPORT REQUIREMENTS

14.5/15.5 Data Collection

2) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method.

14.6/15.6 Quality Controls

3) For multi-mode surveys, detailed description of any data quality issues arising from combining data collected via different modes/instruments.

The proposed revisions the Panel will be asked to comment on involve the following topics:

- Clarify what needs to be stated in the proposal.
- Incorporate high quality benchmark questions in order facilitate detection of mode bias.
- Encourage use of similar modes of administration – i.e. both self-administered or both interviewer-administered – in order to reduce mode effects.
- Pre-testing requirements for multi-mode surveys
- Clarify the reporting requirements in terms of (a) any adjustments made to mitigate mode biases, and (b) the decision to combine versus not combine the data across modes.
- Clarify calculation of outcome rates for multi-mode.

The final question to the Panel will be whether there are any other suggestions for revisions or additions to standards with respect to multi-mode surveys.

F1. *Proposal Documentation* for multi-mode surveys

In Section 1, *Proposal Documentation,* the requirements specific to multi-mode surveys are:

1. PROPOSAL DOCUMENTATION

1.2.4. Description of Data Collection

7) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method.

Other components of Section 1 will ensure that the methodology employed for each mode will be described even though they don't explicitly refer to multi-mode survey designs (e.g. *1.2.2 Sample/Sampling Details,* and other elements in *1.2.4*).

The primary concern associated with multi-mode surveys is the potential for mode bias – that is, getting different response distributions for the same question due specifically to characteristics of the mode. For example, it has been found that social desirability effects tend to be stronger in interviewer-administered surveys (e.g. telephone) than in self-completion surveys (e.g. online). Differences across mode in how questions and response options are presented could potentially cause different response distributions. Mode bias can vary by question – e.g. some questions may be more prone to social desirability effects than others, and some questions may be more similar in design across modes than other questions. These types of mode effects pose challenges for combining or comparing data across modes.

Note that differences in response due to different types of people using the different modes is not a problem, and indeed improving population coverage is a reason to consider doing multi-mode surveys in some circumstances.

The issue is whether and how the proposal documentation requirements need to be elaborated to make it more clear in the proposal that the issue of potential mode bias is recognized and that steps will be taken to address this. The existing requirement only indirectly refers to dealing with the potential for mode bias. Note that Section 4, *Sampling*, is more explicit, but perhaps more can be done in Section 1 of the Standards.

Proposed revisions to 1.2.4 *Description of Data Collection*, and 1.2.7 *Data Analysis*

1. PROPOSAL DOCUMENTATION

1.2.4. Description of Data Collection

7) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method, and a rationale for the specific modes that will be used. Describe the steps that will be taken (a) to reduce the likelihood of mode biases, and (b) to facilitate detection of any mode biases.

1.2.7. Data Analysis

1) Briefly describe how the data will be analyzed in order to address the objectives/research questions, including any special analyses (e.g., segmentation)
   a) For multi-mode surveys, state whether the plan is to report results based on combining the data across modes, or to report results separately by mode.

Questions Posted on the Discussion Board

F1.1 PORD would like to clarify and strengthen the standards for multi-mode surveys. Question series F1 to F6 are to get the Panel's input on a variety of topics related to multi-mode surveys.

Firstly, with regard to *Proposal Documentation*, any suggestions for changes to 1.2.4 *Description of Data Collection*? This adds a requirement to give a rationale for the modes that will be used, and to describe steps that will be taken (a) to reduce the likelihood of mode biases, and (b) to facilitate detection of any mode biases.

F1.2 With regard to *Proposal Documentation*, any suggestions for changes to 1.2.7 *Data Analysis*? This adds a requirement to explain whether or not results will be based on combining data across modes, or if results will be reported separately by mode.

F2. *Sampling Procedures* and *Questionnaire Design* for multi-mode surveys

The existing standard for multi-mode surveys in *Sampling Procedures* is:

4. SAMPLING PROCEDURES

4.5. Multi-Mode Surveys

Multi-mode surveys are ones where different methods of questionnaire administration are used. They will often involve a combination of online and telephone methods, although there are other possibilities (e.g., in-person, mail, fax).

When a survey is conducted using multiple modes of questionnaire administration:

1) The reasons for using a multi-mode rather than a single-mode method must be stated, both in the research proposal and the survey report.
2) When the plan is to combine data collected via different modes in the data analyses, then steps must be taken to ensure as much comparability as possible across the different survey modes in terms of question wording and presentation of response options.
3) Steps must be taken to ensure avoidance of duplicate respondents in different modes. The steps taken, and the results, must be documented.

There is no current standard for *Questionnaire Design* specific to multi-mode surveys.

The intent in the proposed changes below are to (a) distinguish between sampling-related topics and questionnaire-related topics, (b) increase the prominence of the value of using similar modes of survey administration, (c) clarify that one needs to be concerned about mode biases when comparing results by mode as well as when combining data across modes, and (d) to highlight the value that benchmark questions can have for enabling detection of mode biases.

The proposed revisions to *Sampling Procedures* are:

4. SAMPLING PROCEDURES

4.5. Multi-Mode Surveys

Multi-mod

e surveys are ones where different methods of questionnaire administration are used (e.g. some combination of telephone, online, in-person, or mail).

When a survey is conducted using multiple modes of questionnaire administration:

1) The reasons for using a multi-mode rather than a single-mode method must be stated, both in the research proposal and the survey report.
2) The rationale for the specific modes used must be stated, both in the research proposal and the survey report.
    a) The risk of mode biases can be lower if the modes of administration are similar – i.e. both interviewer-administered (e.g. telephone and in-person) or both self-administered (e.g. online and mail).
3) Steps must be taken to ensure avoidance of duplicate respondents in different modes. The steps taken, and the results, must be documented.

The proposed revisions to Section 2 *Questionnaire Design* are as follows -- this would be a separately numbered item in Section 2, and for now the number is shown as "x"; for now we've used the same introductory explanation of multi-mode surveys as in 4.5):

2. QUESTIONNAIRE DESIGN

2.1. Standards

x) Multi-mode surveys are ones where different methods of questionnaire administration are used (e.g., some combination of telephone, online, in-person, or mail).

When a survey is conducted using multiple modes of questionnaire administration:

a) When the plan is to combine or compare data collected via different modes in the data analyses, then steps must be taken to ensure as much comparability as possible across the different survey modes in terms of question wording and presentation of response options.

b) Include questions for which there are high quality data on the population and which therefore can be used as benchmarks for exploring the possibility of mode biases. The required demographic questions can serve this purpose (see Section 1.2.4). Also try to identify any other high quality benchmarks available – whether demographic, behavioural or attitudinal –that might correlate with key measures in the survey. The latter can help better identify any mode biases in the key survey measures.

Questions Posted on the Discussion Board

F2.1 The Background proposes changes to Section 4 *Sampling Procedures* and Section 2 *Questionnaire Design*, with the intent to (a) distinguish between sampling-related topics and questionnaire-related topics, (b) increase the prominence of the value of using similar modes of survey administration, (c) clarify that one needs to be concerned about mode biases when comparing results by mode as well as when combining data across modes, and (d) to highlight the value that benchmark questions can have for enabling detection of mode biases.

Do you have suggested changes to the proposed revision of Section 4.5 *Multi-mode surveys*?

F2.2 Do you have any suggested changes to the proposed revision of Section 2.1, which is standards for *Questionnaire Design?*

F3. Section 3 *Pre-Testing* for multi-mode surveys

The current Section 3 standard for pre-testing does not make any specific references to multi-mode surveys.

For reference, Section 3 *Pre-Testing* (online version) is:

3. PRE-TESTING

3.1. Standards

1) Pre-testing of all components of a new or revised survey questionnaire that may influence data quality and respondent behaviour is required. This includes the online appearance and functionality of the questionnaire.
2) The client must be given the opportunity to test and approve the online survey prior to launch.
3) Pre-testing must include probing that invites participants recruited for this purpose to provide input about their comprehension of and reaction to the questions. For example, a short series of questions could be included at the end of the pre-test survey. Researchers and clients must agree in advance as to whether probing will take place during or after administering the survey. If requested by the client a cognitive pre-test must be conducted.
4) To help ensure questionnaire effectiveness with subgroups where there is reason for concern (e.g., due to language, age, level of education, etc.), the socio- demographic characteristics of the targeted participants must be approved by the client before recruiting begins.
5) A minimum of 10 pre-test surveys are to be completed in each language in which the final survey will be fielded. An exception could be projects with small survey populations, such as a client-satisfaction survey of a small client base. In such cases the researcher must, in consultation with the client, take steps to ensure that the smaller number of pre-tests are sufficient to guarantee questionnaire quality. For example, a cognitive pre-test may be warranted.
6) Pre-test completions shall not be included in the final dataset. An exception could be projects with:
   a) hard-to-reach target groups, or
   b) when no changes are made to the questionnaire.
7) Documentation of the pre-test(s) must be provided to the client before the questionnaire is finalized. The documentation must include (at minimum):
   a) a description of the pre-test approach and number of interviews completed;
   b) findings and any resulting modifications;
   c) average survey completion time;
   d) a statement of whether or not pre-test cases will be retained in the final data set.

   The final research report must include this same information.

Questions Posted on the Discussion Board

F3.1 Section 3.1.3 of *Pre-Testing* stipulates a minimum 10 English and 10 French pre-test interviews, and there is no reference to multi-mode surveys. In the case of a multi-mode survey, should there be a requirement for a minimum number of pre-test interviews in English and French for each mode? If so, what should be the minimum number of interviews per mode in English and French?

F4. Section 8 *Outcome Rates* for multi-mode surveys

Currently in Section 8 *Outcome* Rates there is no standard for how to report outcome rates for a multi-mode survey.

The following proposal is based on general recommendations in the AAPOR document, *Standard Definitions: Final dispositions of case codes and outcome rates for surveys*[51]. A more detailed discussion can be found in the Sage Research report, *Best Practices for Improving Cooperation for Online Surveys*[52].

The proposed text below would be part of the "definitional" part of Section 8 (8.1 to 8.4), and before the "mandatory requirements" part (8.5). For now it is labeled "8.x".

8. OUTCOME RATES

8.x. Multi-mode surveys

1) The general principles for calculating the response rate for mixed-mode surveys are (a) calculate a response rate for each mode or stage of sampling, (b) use the same general response rate formula for each mode or stage of sampling, and (c) calculate a single aggregated response rate. The specific details of how this is done will be adapted to the particular multi-mode survey design.

Questions Posted on the Discussion Board

F4.1 Currently in Section 8 *Outcome Rates* there is no standard for how to report outcome rates for a multi-mode survey. The Background proposes a standard consisting of three principles: (a) calculate a response rate for each mode or stage of sampling, (b) use the same general response rate formula for each mode or stage of sampling, and (c) calculate a single aggregated response rate. Do you have any suggestions for changes to the proposed standard?

F5. Section 14 (online)/15 (telephone) *Mandatory Survey Report Requirements* for multi-mode surveys

The existing standards referring to multi-mode surveys are:

14 (online)/15 (telephone). MANDATORY SURVEY REPORT REQUIREMENTS

14.5/15.5 Data Collection

2) For multi-mode surveys, provide a rationale for using a multi-mode rather than a single-mode method.

14.6/15.6 Quality Controls

3) For multi-mode surveys, detailed description of any data quality issues arising from combining data collected via different modes/instruments.

14.5/15.5 # 2 will be updated based on the Panel discussion of 1.2.4 in *Proposal Documentation*, where this same language was used.

[51] AAPOR. (2011). *Standard Definitions: Final dispositions of case codes and outcome rates for surveys. 7th Edition.* Deerfield, IL: The American Association for Public Opinion Research.

[52] Sage Research Corporation. (2015). *Best Practices for Improving Cooperation for Online Surveys.* Prepared for the Public Opinion Research Directorate, Government of Canada.

The intent of the proposed revisions to 14.6/15.6 are to (a) ensure there is clarity as to decisions made about combining or not combining data across modes, and (b) to require descriptions of any adjustments made to the data to mitigate mode effects.

Proposed revision to 14.6/15.6 *Quality Controls*

14.6/15.6 Quality Controls

3) For multi-mode surveys, detailed description of any data quality issues arising from combining data collected via different modes/instruments. The discussion must also include:
   a) Rationale for decisions about combining or not combining data from different modes in reporting the results.
   a) Description of any steps taken to mitigate mode effects in the survey results.

Questions Posted on the Discussion Board

F5.1 The existing standard in the *Quality Controls* section of *Mandatory Reporting Requirements* addresses *"description of any data quality issues arising from combining data collected via different modes/instruments."* The Background proposes an additional requirement: *Description of any steps taken to mitigate mode effects in the survey results.*

Is this additional necessary/useful? Any changes to the wording?

F6. Any other revisions to standards recommended for multi-mode surveys?

Questions Posted on the Discussion Board

F6.1 Do you have suggestions for any other revisions to the Standards to address issues associated with multi-mode surveys?

G. Incentives in Surveys of Children, Young People or Vulnerable Respondents

Section 6 *Data Collection from Children, Young People or Vulnerable Respondents* does not make any reference to whether or how incentives are used for this survey population.

Section 7.5 (telephone)/7.6 (online) does not refer to this population either:

7.5/6. Incentives/Honoraria

1) The details of any incentives/honoraria to be used for an telephone/online survey must be provided in both the proposal and report documentation, including:
   a) the type of incentive/honoraria (e.g., monetary, non-monetary);
   b) the nature of the incentive (e.g., for monetary, prize, points, donations, direct payments);
   c) the estimated dollar value of the incentives to be disbursed.

GC's *Qualitative Research* standards contain one standard pertaining to incentives for this target population:

> **5. Participant Recruiting**
>
> **5.4. Incentives**
>
> 7) Ethical considerations should be taken into account especially when providing incentives to youth, or vulnerable groups (e.g. payment is not coercive, or exposes young or vulnerable persons to a risk that they would otherwise have avoided).

PORD received some input from MRIA. MRIA posed the following question to some members (note that it's not clear if the context of the question was survey research, or if it included qualitative research as well): *Information gathered from audit of the MRIA Gold Seal members and speaking with other members, shows that contact with children and young people are through their parents and guardians. Please provide your comments on how your firm offers incentives to children and young people.* PORD relayed the following response from MRIA:

> Perform household level research, so incentives are offered to all members of the household regardless of age.
>
> The incentive is always in cash - recognizing that young people may not be able to cash a cheque themselves.
>
> Some members offer the incentives to the parent/guardian on behalf of the child.
>
> Gift cards related to gaming are offered by some members.

PORD would like the Panel to consider whether the Telephone and Online Standards should have guidance on incentives for surveys of children, young people or vulnerable respondents.

- The Standards define children and young people as follows:
  - Children: Under the age of 13
  - Young people: 13-15

  People 16 and over do not require the consent of a parent or responsible adult (guardian, etc.) before being invited to participate in a survey.

If Panel members think there should be guidance for this survey population, the following is an addition to 7.5/6 *Incentives/Honoraria* that is a starting point for the Panel's consideration:

> **7.5/6. Incentives/Honoraria**
>
> 1) The details of any incentives/honoraria to be used for an telephone/online survey must be provided in both the proposal and report documentation, including:
>
> a) the type of incentive/honoraria (e.g., monetary, non-monetary);
>
> b) the nature of the incentive (e.g., for monetary, prize, points, donations, direct payments);
>
> c) the estimated dollar value of the incentives to be disbursed.
>
> 2) When survey respondents are children, young people or vulnerable individuals, and an incentive is being offered:
>
> a) Decide in advance who will receive the incentive – the parent or responsible adult (guardian, etc.), the respondent, or if both will receive an incentive.

b) The parent or responsible adult must agree to the incentive, regardless of who is receiving the incentive.

c) Ethical considerations should be taken into account when providing incentives to children, youth, or vulnerable groups (e.g. payment is not coercive, or exposes young or vulnerable persons to a risk that they would otherwise have avoided).

Questions Posted on the Discussion Board

G1. There are currently no standards pertaining specifically to incentives when the respondents are children, young people or vulnerable individuals. Should there be any standards/guidelines on this topic? If so, what should these be? The Background gives a possible standard you might want to edit.

Appendix C: Background and Questions – Discussion Board #3

DISCUSSION BOARD #3: BACKGROUND AND QUESTION GUIDE
(Start date: June 4; End date: June 15)

Topics:

The following contains the general questions for the group discussion. More detailed questions may be developed for the four topics covered in this first Discussion Board as it progresses and/or based on the views of Panel members.

Please note the following:

1) **You will need to refer to this Agenda for the detailed commentary on the questions shown on the Discussion Board.** The Discussion Board displays only the questions. This Agenda also provides, for example:
   -- specific excerpts from relevant sources
   -- references to the information available and the specific pages in the *Literature Review* report where this information can be found

   This type of detail has **not** been included in the Discussion Board questions.

2) Please answer **all** the relevant questions; it is important to know the point of view of each member of the Advisory Panel.

3) Wherever possible, we need to have as detailed answers as possible, particularly in areas where there is possible disagreement between Panel members on either a principle stated, or the language used to express the principle.

4) We need all members of the panel to begin posting answers to the questions as soon as possible and shortly after the Discussion Board is up and running. Otherwise, there will be little

if any time left to debate issues, to try to get clarification or more detail from the Panel, or to come to some agreement on issues where Panel members have different views.

It is the mandate of the Advisory Panel ***to reach consensus where possible even though it is not an essential outcome of deliberations***. Where there is initial disagreement, please talk to one another to see if a consensus can be found. Think of lack of consensus as a last resort.

5) Just to remind you about the terminology being used:

Standards	Practices which are ***requirements*** for all quantitative research conducted by the Government of Canada
Guidelines	Practices which are ***recommended,*** but would not be requirements; that is, known good practices or criteria that serve as a checklist to ensure quality research but are not necessarily applied to every study

Wherever possible for the sake of simplicity, the standards or guidelines should be the same for Government of Canada POR telephone and online surveys.

6) PORD asks that the following criteria be applied:

In general, we wish to adhere to the following criteria when revising/developing Standards:

- *Be measurable*
- *Not conflict with other GC departments (e.g., TBS) or related policies or standards (e.g., IT, legal)*
- *Avoid material that will quickly become out-of-date*
- *Not overly prescriptive*
- *Minimize web links in the content*

7) You will see various citations in the footnotes of this Background. We have tried to make the Background "self-sufficient" so that it is not necessary to read the cited articles. However, if you would like to get any of these documents, let us know which ones you would like and we'll email them to you.

8) Panel members will need to refer to the following document, and so is being provided to members:
   - MRIA's Appendix C – *Guideline on Social Media Research*

A. Privacy and Security of Data

A1. Data Breaches

A data breach is the loss of or unauthorized access to/disclosure of personal or organizational information.

The current standards require taking steps to protect against data breaches. The objective for the Panel is to identify any revisions or additions to the standards, and/or any guidelines that should be included.

The current standards are:

13.2 (Online)/14.2 (Telephone). Protection of Data/Servers

1) Protection against illegal or unsanctioned access: Researchers must use up-to-date technologies to protect survey data collected or stored on Web sites or servers against illegal or unsanctioned access by third parties (i.e. “hacking”). The researcher must also control access to all databases on which any data relating to the survey is stored so that only individuals with the appropriate security clearance are able to access the database, either by using a password or other form of access restriction (such as biometric controls).
3) The researcher must not subcontract (including to an affiliate) any function that involves providing a subcontractor with access to any data relating to the survey unless the client first consents in writing.
5) Protection against physical damage to servers: Researchers must also put in place measures to ensure the “physical” security of data and servers.

13.3/14.3 Temporary Storage of Data on Servers

1) If the temporary storage of data collected takes place on a server that is operated by another provider, the researcher must place the provider under the obligation to take the necessary steps to ensure that the requirements described in subsection 13.2. are met. Temporary storage of the collected data on the server must be terminated at the earliest possible time.

13.6/14.5 In the Event of Any Data Breach

1) In the event of any data breach, the client must be informed immediately and provided with details about both the nature and the extent of the data breach.

The literature review commissioned by PORD (see pp. 28-30) cites the following as a framework for considering data protection:

- **Be aware of the data the organization has** – Know exactly what kind of data the organization has, where/how it is stored, as well as where/when it is collected, and who has access. When organizations have a clear understanding of the data, they can identify the type of data that would require a unique protection system and they can adopt or develop approaches to safeguard these data.
- **Be aware of the organization’s vulnerabilities** – Risk and vulnerability assessments help to ensure that threats to privacy are identified and addressed. The vulnerabilities organizations should be aware of include: third-party activities involving the organization and data.
- **Limit the information collected and the length of time the information is retained** – It is not only important to know why you collect the information, but organizations should know why they are holding this information.
- **Clearly define policies and procedures about the secure destruction of information** – Improper disposal of the information can lead to data leakage.
- **Train employees** – Policies can only be effective when those responsible for implementing and abiding by them are aware of what they contain, why they exist, and the consequences of neglecting their responsibilities.
- **Maintain up-to-date software and safeguards** – Establish routine and documented steps to ensure security-related updates are applied in a timely manner, and that software no longer in use is removed from the system.
- **Implement and monitor intrusion prevention and detection systems** – Measures such as intrusion detection systems, firewalls and audit logs can help to identify and respond to privacy breaches before they escalate.
- **Encrypt laptops, USB keys and other portable media**.

PORD would like to know whether this framework suggests any revisions to the current standards cited above.

Questions Posted on the Discussion Board

A1.1 The Background lists the current standards pertinent to protection against data breaches, and cites a data protection framework in the literature review commissioned by PORD.

Does the framework in the literature review suggest any revisions/additions to the current standards, or any guidelines, with respect to preventing data breaches? Or, are the existing standards sufficient? For example, should anything more be stated with respect to:

- Destruction of information

  Note: Section 13.1.1/14.1.1 deals with retention of technical data but does not deal with either (a) destruction of data in general or (b) destruction of information in the event of a data breach.

- Encryption

A2. Passive Data Collection in Online Surveys

Online and mobile methodologies create possibilities for collecting various types of personal data "passively", that is without direct interaction with respondents. The issue to be considered is, in the context of surveys, what passive data collection is allowed and under what circumstances is it allowed?

The ESOMAR/GRBN *Guideline on Mobile Research*[53] states:

> Mobile applications are capable of collecting a broad range of personal data without direct interaction with data subjects. Examples include web use and browsing history, app usage statistics, loyalty card data, geolocation, social media data, data from wearables and IoT and other data generated by or obtained from mobile devices.

At least some of these capabilities can also apply to online surveys where the device used by the respondent is a computer.

The *Guideline* notes an important exception to what constitutes "personal" data involving certain device characteristics:

> While it is possible to passively detect the type of device a data subject is using, this is not personal data as long as the purpose is to optimize app performance and survey rendering.

The existing Standards address relevant general principles of protection of respondent anonymity and confidentiality, and address passive data collection to some extent.

[53] ESOMAR and GRBN. (2017). *Global Guideline on Mobile Research*

- Section 5.1.4, *Protection of Anonymity and Confidentiality*, states the general principle: *The anonymity of respondents must always be preserved unless they have given their informed and explicit consent to the contrary.*

> **5.1.4. Protection of Anonymity and Confidentiality**
>
> 1) The anonymity of respondents must always be preserved unless they have given their informed and explicit consent to the contrary. If these respondents have given informed consent for data to be passed on in a form which allows them to be personally identified, the researcher must ensure that the information will be used for research purposes only, OR, if requested by the respondent, to resolve a customer complaint. The same holds true when respondents' answers are collated, with their informed consent, with pre-existing data that allows such identification (e.g., with an administrative data base). Moreover, such personally identifiable information must not be used for any purpose unrelated to the current study, such as direct marketing, list-building, credit rating, fund-raising, or any marketing activities directed at those individual respondents.

  Section 5.1.4 establishes that any personal data collected passively must be handled in a way that protects the anonymity and confidentiality of the survey respondents.

- Section 5.3, *Privacy Issues Specific to Online Survey research*, states a requirement to have an accessible policy statement *"concerning the use of cookies, log files and, if applicable, software."* This ensures respondents are informed of certain types – but not all types – of passive data collection.

> **5.3 Privacy Issues Specific to Online Survey Research (Online)**
>
> 1) Researchers must have a readily accessible policy statement concerning the use of cookies, log files and, if applicable, software. This statement may be either included in their privacy policy or it may appear in a separate document. Software must not be installed on respondents' computers without their knowledge or consent. In addition, respondents must be able to remove the researcher's software easily from their machines (e.g., for Windows users, the software must appear in the Add/Remove Programs folder in their Control Panel).
> 2) Any links to data protection, privacy policy or cookie policy statements must be given at the start of the questionnaire.

- Section 7.2, *Data Collection and Recruitment Techniques* (Online Standards), addresses passive data collection with its reference to forbidding "surreptitious" or "unsolicited" data collection – terms which on the face of it would appear to refer to passive data collection generally. The principle is that passive data collection can only be done with the respondent's awareness and presumably consent (although the standard doesn't say "consent"):

> **7.2. Data Collection and Recruitment Techniques**
>
> 1) Researchers must not make use of surreptitious, misleading or unsolicited data collection or recruitment techniques – including using spambots, spiders, sniffers or other 'agents' that collect personal information without the respondent's explicit awareness.

It may be that the above standards are sufficient to address issues associated with passive data collection in surveys, and the Panel will be asked if this is the case.

Alternatively, it may be that the existing standards are not considered to be sufficiently explicit. To aid in Panel consideration of this topic, we took a crack at reformulating standard 7.2:

Possible revision of 7.2. Data Collection and Recruitment Techniques

1) Passive data collection refers to collecting personal information about respondents without direct interaction with them.

   Researchers must not make use of data collection or recruitment techniques using passive data collection methods unless survey respondents or potential respondents have first given informed consent, or unless collecting the information is legally permissible or is permissible under the Terms of Use of the website, service or application from which the data are sourced.

   Examples of passive data collection of personal information include, but are not limited to, web use and browsing history, app usage statistics, geolocation, personally identifiable biometric data, social media data, data from wearables and IoT (internet of things), and other data generated by and obtained from respondents' mobile devices or computers.

   Passive detection of the type of device a respondent is using is not personal data as long as the purpose is to optimize app performance and survey rendering.

Some things considered when formulating this possible revision to 7.2:

- The Panel is reviewing standards for online and telephone surveys, so the language tries to emphasize the survey context for the standard.
- The original 7.2 referred to both "data collection and recruitment techniques", so both have been retained.
- The proposed revision includes a definition of "passive data collection" of personal information.
- ESOMAR/GRBN standards make references to exceptions where the data collection is "legally permissible" or is allowed under "Terms of Use." This has been incorporated into the proposed revision of 7.2.
  - -- ESOMAR/GRBN *Guideline on Mobile Research*, Section 3.4 on *Passive data collection*: *"When it is not possible to obtain consent (such as measuring traffic to a website), researchers must have legally permissible grounds to collect the data…"* (p. 10).
  - -- ESOMAR/GRBN *Global Guideline – Online Research*, Section 3.2.1 *Passive data*:

    > Where personal data are collected from public spaces such as websites or social media sites, consent must be obtained directly or explicitly provided for in the Terms of Use (ToU) policy of the platform. This does not apply to publication in social media that includes the author's name, which implies a diminished expectation of privacy. (p.10)

    Note: In our view the last sentence referring to not applying to *"publication in social media that includes the author's name"* does not apply to survey research. Even if there is passive collection of such information, it forms part of a data record that includes responses to a survey questionnaire. Standard 5.1.4 ensuring anonymity and confidentiality therefore takes precedence.
- Another reason for the inclusion of "legally permissible" and "Terms of Use" involves surveys done using a panel or a MROC (market research online community). If the survey incorporates information available on panel/community members, is that considered to be passive data collection? We're not sure, so let us know what you think. If it is within the scope of "passive data collection", does the reference to the "Terms of Use" deal with this?

  Also, does the proposed phrase, "*website, service or application"*, clearly include panels/communities, or does it need revision?

Note: By including panels (and perhaps MROCs) within the scope of "passive data collection of personal information", should the proposed standard also be incorporated into the Telephone Standards? Most of the possibilities for passive data collection involve apply to online surveys. Are there passive data collection issues with telephone surveys?

- The ESOMAR/GRBN *Guideline on Mobile Research*, Section 3.4 *Passive data collection*, includes a subsection 3.4.1 *Biometric data*. The examples they cite:

  > For example, facial coding involves recording a data subject's face as he or she completes a survey or similar task. Eye tracking, virtual reality headsets and other wearable devices may be used in a similar way. (p.10)

  To address this, the proposed revision to 7.2 includes *"personally identifiable biometric data"* (not all biometric data is personally identifiable; for example, an eye-tracking gaze pattern superimposed on the object being looked at does not identify the individual).

- The proposed revision explicitly excludes device type detection from the informed consent requirement. It uses the language in the ESOMAR/GRBN *Guideline on Mobile Research* (the footnote on p.10).

Questions Posted on the Discussion Board

A2.1 Online and mobile methodologies create possibilities for collecting various types of personal data "passively", that is without direct interaction with respondents.

Do the current Standards, as outlined in the Background, adequately address passive data collection of personal information done in conjunction with a survey?

To facilitate discussion, the Background gives an alternative to Standard 7.2 *Data collection and recruitment techniques* that deals with passive data collection in a different way. What do you think of this alternative? If you think something like this is needed, please revise as you see fit.

For ease of reference, the alternative version is:

Possible revision of 7.2. Data Collection and Recruitment Techniques

1) Passive data collection refers to collecting personal information about respondents without direct interaction with them.

   Researchers must not make use of data collection or recruitment techniques using passive data collection methods unless survey respondents or potential respondents have first given informed consent, or unless collecting the information is legally permissible or is permissible under the Terms of Use of the website, service or application from which the data are sourced.

   Examples of passive data collection of personal information include, but are not limited to, web use and browsing history, app usage statistics, geolocation, personally identifiable biometric data, social media data, data from wearables and IoT (internet of things), and other data generated by and obtained from respondents' mobile devices or computers.

   Passive detection of the type of device a respondent is using is not personal data as long as the purpose is to optimize app performance and survey rendering.

A3. Photographs and Recordings

The Online and Telephone Survey Standards do not currently have any standards pertaining specifically to respondent photographs, videos or audio recordings.

The ESOMAR/GRBN *Guideline on Mobile Research* Section 3.4.2 states the following about *Photographs and recordings:*

> Photographs, video and audio recordings are considered to be personal data and therefore must be gathered, processed and stored as such. They can only be shared with a client if the data subject gives his or her prior consent with knowledge of the specific purpose for which it will be used. When potentially identifying information has been removed (such as through pixelisation or voice modification technology) so that it is no longer considered personal data it can be shared with a client provided the client agrees to make no attempt to identify the individual.
>
> Researchers must not instruct data subjects (or those that may be acting as data collectors) to engage in surveillance of individuals or public places. Data subjects should be given specific limited tasks (e.g. capturing interactions with friends with their consent, or images of objects or displays) that do not involve monitoring a particular area where personal data would be captured without the consent of the individuals present. When recorded observation of a location is undertaken, clear and legible signs indicating that the area is under observation along with the contact details for the researcher or research organization performing the research should be posted and images of individuals must be pixelated or deleted as soon as possible. Cameras should be situated so that they monitor only the areas intended for observation. (pp. 10-11)

The following is a slightly revised version of the ESOMAR/GRBN guideline that emphasizes the survey context (as the Panel is only considering revisions to the Telephone and Online Survey Standards):

> **Revised to emphasize survey context:** Photographs, video and audio recordings from survey respondents are considered to be personal data and therefore must be gathered, processed and stored as such. They can only be shared with a client if the respondent gives his or her prior consent with knowledge of the specific purpose for which it will be used. When potentially identifying information has been removed (such as through pixelisation or voice modification technology) so that it is no longer considered personal data it can be shared with a client provided the client agrees to make no attempt to identify the individual.
>
> Researchers must not instruct survey respondents to engage in surveillance of individuals or public places. Respondents should be given specific limited tasks (e.g. capturing interactions with friends with their consent, or images of objects or displays) that do not involve monitoring a particular area where personal data would be captured without the consent of the individuals present. When recorded observation of a location is undertaken, clear and legible signs indicating that the area is under observation along with the contact details for the researcher or research organization performing the research should be posted and images of individuals must be pixelated or deleted as soon as possible. Cameras should be situated so that they monitor only the areas intended for observation. (pp. 10-11)

Questions Posted on the Discussion Board

A3.1 The Background gives a standard pertaining to photographs, videos and recordings closely based on an ESOMAR/GRBN standard.

The Online and Telephone Survey Standards do not currently have any standards pertaining specifically to respondent photographs, videos or audio recordings. Do the Online and Telephone Survey Standards need a standard on this topic?

If a standard is needed, what do you think of the modified ESOMAR/GRBN standard in the Background? Please revise as you see fit.

A4. Telephone Surveys – Sensitivity to Setting

PORD has requested the Panel consider whether there needs to be an addition to the Telephone Standards related to sensitivity to the respondent's setting. In this regard, there are two relevant guidelines in the ESOMAR/GRBN *Guideline on Mobile Research* (p. 7):

> **3.1.1 Safety**
> When calling mobile phones researchers may sometimes contact potential data subjects who are engaged in an activity or in a setting not normally encountered in fixed-line calling. This might include driving a vehicle, operating machinery or walking in a public space. The researcher should confirm whether the individual is in a situation where it is legal, safe and convenient to take the call. If the researcher does not receive confirmation, then the call should be terminated while allowing the possibility of making further attempts at another time.
>
> **3.1.2 Confidentiality and sensitive data**
>
> A researcher might contact a potential data subject who is engaged in an activity or situation where others may overhear the call. In this case, the researcher must consider the nature of the research content in light of the possibility that the data subject might be overheard and personal information or behaviour inadvertently disclosed or responses modified as a result of their situation. If appropriate, the call should be rescheduled to another time or location when confidentiality will not to be compromised.

Note that the ESOMAR/GRBN 3.1.1 above is specific to mobile phones, but 3.1.2 could apply to either mobile or fixed-location phones.

The current Telephone Standards, in Section 5.2 #1 *Avoidance of Harassment*, has a standard focused on sensitivity of the survey subject matter, but it does not directly address issues caused by the setting of the interview:

> **5.2. Avoidance of Harassment**
>
> 1) The researcher must take all reasonable steps to ensure that respondents are not in any way hindered or embarrassed by any interview, and that they are not in any way adversely affected as a result of it. Researchers must address sensitive subject matter in a way that will minimize the discomfort and apprehension of both respondents and interviewers.

Questions Posted on the Discussion Board

A4.1 Because respondents are increasingly likely to answer calls using a mobile phone, there can be issues with them using the phone in problematic settings (e.g. driving, walking in a public space). On both mobile phones and fixed-location phones, they may be in a setting where they can be overheard.

Does there need to be a standard (or guideline) for the interviewer to confirm the respondent is in a location where they are comfortable taking the call? Or, is it reasonable that it is the responsibility of the respondent to raise this if they have an issue, and not have a standard (guideline)?

If a standard or guideline is needed, how would you word it?

A starting point for language: The interviewer must (should?) confirm with a respondent that the respondent is in a location where they are comfortable doing the interview.

A5. Cloud Storage

The current Standards require that survey data must be stored in Canada:

13.2. Protection of Data/Servers

2) Protection against legally-sanctioned access: Because some jurisdictions allow their authorities, under certain circumstances, to access all data stored on servers located in that jurisdiction (e.g., in the United States under provisions of the *Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act,* known as USA PATRIOT Act: see the Treasury Board Secretariat 's overview), the researcher must ensure that all the databases containing any information related to the survey are stored on **servers and back- up servers** located solely in Canada.

   a) If the client has first consented in writing, these servers may be located in another country where:
      i. equivalent protections are given to personal information as in Canada under legislation such as the *Privacy Act*, R.S. 1985, c. P-21, and the *Personal Information Protection and Electronic Documents Act*, S.C. 2000, c. 5, and under any applicable policies of the Government of Canada; and
      ii. the laws do not allow the government of that country or any other entity or person to seek or obtain the right to view or copy any information relating to the survey without first obtaining the client's written consent.

   b) In connection with giving consent to locating a database in another country, the client may, at their option, require the researcher to provide a legal opinion (from a lawyer qualified in the foreign country) that the laws in that country meet the requirements of paragraph 2) a), above, or may require the researcher to pay for the Government of Canada to obtain such a legal opinion. The Government of Canada has the right to reject any request to store survey data in a country other than Canada if there is any reason to be concerned about the security, privacy, or integrity of the data. The Government of Canada may also require that any data sent or processed outside of Canada be encrypted with Government of Canada-approved cryptography and that the private key required to decrypt the data be kept in Canada in accordance with key management and storage processes approved by the Government of Canada.

c) The researcher must ensure that all servers, including back-up servers, on which any data relating to the survey is stored are physically and logically independent (meaning there is no direct or indirect connection of any kind) from all other databases, unless those databases are located in Canada (or in another country approved by the client under paragraph 2) a)) and otherwise meet the requirements of this section.

d) The researcher must ensure that all data relating to the survey is processed only in Canada or in another country approved by the client under paragraph 2) a).

e) The researcher must ensure that all domestic network traffic (meaning traffic or transmissions initiated in one part of Canada to a destination or individual located in another part of Canada) is routed exclusively through Canada, unless the client has first consented in writing to an alternate route. The client will only consider requests to route domestic traffic through another country that meets the requirements of paragraph 2) a).

Questions Posted on the Discussion Board

A5.1 The current Standards require that survey data must be stored in Canada. PORD would like to know if the Panel thinks any other standards are required with respect to cloud-based storage, either in terms of location of servers/back-up servers or any other aspects of data security specific to cloud-based storage. Any suggestions?

B. Accessibility and Literacy

B1. Should there be a general statement in the Standards promoting accessibility, usability, inclusion and literacy?

The Online and Telephone Standards do not contain any standards or guidelines pertaining to accessibility or literacy.

The "B" series of questions address several aspects of this topic. **We suggest you look over the whole series of questions before answering the question in this section (B1).**

PORD refers to the Web Accessibility Initiative (WAI) home page (https://www.w3.org/WAI/intro/usable) for definitions of accessibility, usability and inclusion:

> **Accessibility:** *Accessibility* addresses discriminatory aspects related to equivalent user experience for people with disabilities, including people with age-related impairments. For the web, accessibility means that people with disabilities can perceive, understand, navigate, and interact with websites and tools, and that they can contribute equally without barriers.
>
> **Usability:** *Usability* and *user experience design* is about designing products to be effective, efficient, and satisfying. Specifically, ISO defines usability as the "extent to which a product can be used by specified users to achieve specified goals effectively, efficiently and with satisfaction in a specified context of use."
>
> **Inclusion:** *Inclusive design*, *universal design*, and *design for all* involves designing products, such as websites, to be usable by everyone to the greatest extent possible, without the need for adaptation. Inclusion addresses a broad range of issues including access to and quality of hardware, software,

and Internet connectivity; computer literacy and skills; economic situation; education; geographic location; and language — as well as age and disability.

PORD has asked the Panel to consider: *[Does the Panel] see a way of condensing the discussion around accessibility, usability and inclusion with concepts of literacy…?*

Questions Posted on the Discussion Board

B1.1 Should a statement be put into the Standards about the importance of the principles of accessibility – including literacy considerations, usability and inclusion?

If yes:

- Please suggest a wording.
- Would this be a standard (required), a guideline (recommended), or something else – such as a "statement of principles"?
- Would it replace any of the more specific recommendations you make in the subsequent questions in this series?

B2. Accessibility for Online Surveys

The current Online Standards do not contain any standards or guidelines pertaining to the accessibility of online surveys to people with disabilities or to other people who may have difficulty completing an online survey.

PORD has communicated to us that: *PORD has been advised that Treasury Board Secretariat is working on a proposed policy for accessibility standards specific to all devices used to access online surveys. The results of this development work will probably be a year in the making.*

When the Treasury Board Secretariat policy is finalized, it will take precedence.

Note that the Panel has previously considered the question of whether there should be guidelines for mobile-friendly questionnaires. Depending on Panel decisions there, this could address some accessibility issues.

On pages 25-28 of the literature review commissioned by PORD is a section on *Accessibility*. In this section, there are three lists of guidelines from various sources. In the literature review, they are associated with *making mobile surveys* accessible, but you will see that many of the items pertain to online surveys generally.

The three lists are as follows (see the literature review for full context):

Page 26: Web Content Accessibility Guidelines:

- providing text alternatives for any non-text content which allows it to be changed into other forms, such as large print, braille, or speech;

- creating content that can be presented in different ways without losing information or structure;
- separating the foreground from the background so it is easier to see and hear content;
- making all functionality available through a keyboard;
- making pages appear and operate in predictable ways;
- maximizing compatibility with user agents, including assistive technologies, such as screen magnifiers, visual reading assistants, screen readers with synthesized speech or braille, text-to-speech software, speech recognition software, alternative keyboards and alternative pointing devices

Pages 26-27: W3C's Web Accessibility Initiative "that could apply to survey research":

- Avoiding features that require the user to input text rather than selecting from values. In the context of survey research, this means using open-ended questions sparingly.
- Minimizing the amount of information on each page because the screens of mobile devices are typically smaller than those of desktop and laptop computers. For survey research, this means one question per screen. Large grid style questions, therefore, are best avoided.
- Ensuring form fields are presented below rather than beside their labels because screens are smaller, and scrolling side-to-side is not as convenient as scrolling down. In the context of survey research and design, this means positioning a field to enter data, such as a text box for an open-ended question or a box to enter date of birth, underneath the question.
- Ensuring text is resizable up to 200 percent, without the use of assistive technology.
- Avoiding high-resolution images that use up bandwidth (which can be expensive depending on a user's data plan and/or time consuming to access in an area where high speed internet access is not accessible). This could apply to the images used to brand a mobile survey or that are presented to respondents for reaction (e.g., ad testing).

Page 27 "commercial tools" list:

- JavaScript-based questions are best avoided because they tend to not work well with screen readers. This means some of the more interactive and visually engaging types of questions, depending on the survey tool, should not be used for surveys that must be accessible. This includes, for example, drag and drop questions (often used for ranking items), slider questions (often used for rating scale questions), heat map questions, and questions designed to sum the numeric data entered by respondents.
- Make questions as concise as possible. For respondents using screen readers, an excessively wordy question will be tedious to listen to (just as it would be when read by a telephone interviewer). If the question is too long, or there are too many response options, the respondent may have difficulty answering the question and/or provide an unintended response.
- Use grid style questions sparingly. In fact, rather than use a table/matrix to concisely present items/statements for a respondent to rate, it is best to opt for a survey with more screens/pages, and give each item/statement to be rated its own screen/page.
- Make sure warning messages in a survey can be read by screen readers.
- Make the layout as simple as possible and use one screen/page per question, which is more user-friendly for anyone using a screen reader.

The question to the Panel is whether any of the above items suggest standards or guidelines to include in the update of the Online Standards. Some thoughts about how to approach this question:

- As noted earlier, the current Standards do not have any accessibility standards. It may be that you don't think any of the items above should be included in the update as either standards or guidelines. That's OK – and if this applies, please just give your reasons.
- You don't need to comment on every one of these items. Just pick out the ones, if any, that you think should be incorporated into the updated Online Standards.
- A few of these may have been discussed in the previous section on features of a mobile-friendly questionnaire (e.g. use of grids; number of questions per screen). There is no need to repeat that discussion, unless there's something you want to add.
- At this stage, we don't necessarily need to recommend what section of the Standards an item might go into. The relevant section will depend on what the item is and what is being said about it. The interest at this point is more on what, if any, of the items should be considered for inclusion.

Questions Posted on the Discussion Board

B2 There are currently no accessibility standards in the Online Standards. The Background lists various accessibility-related practices listed in the literature review commissioned by PORD, and some factors to consider when evaluating these items. Are there any that you think should be incorporated into the Online Standards update? Please explain your selections and any suggested wording modifications.

B3. Literacy and Online Surveys

The current Online Standards do not contain any standards or guidelines pertaining to literacy.

The literature review commission by PORD states the following about literacy and survey accessibility:

> **pp.27-28:** Usability research provides some guidance on how to design online surveys that are accessible to people with low literacy skills. For example, make questions short in length and use plain language (i.e., avoid the use of jargon, three or more nouns in a row, the passive voice, and verb-noun phrases). This is a best practice for survey questionnaire design in general, but one that is particularly salient when the target population has low literacy skills.
>
> In addition to language, online surveys allow survey researchers to use design elements to help the survey respondent understand and complete the survey. For example, the following style and design conventions can be helpful: using a larger font size, having a clean, simple design (one that is not encumbered by too many colours or graphics), and using bolded font to draw attention to "working" words that convey the meaning of the question or instructions that are important for a respondent (and facilitate his/her reading of the information).
>
> **p.28:** Finally, when it comes to literacy, questionnaires should use plain language to the extent possible and warranted (for a survey of the general public this would be imperative,

but for a survey of a specialized population, this may not be possible or desirable depending on the subject matter).

In a footnote, the literature review refers to The Canadian Style Guide, Section 13, *Plain Language*, which gives a variety of guidelines (http://www.btb.termiumplus.gc.ca/tcdnstyl-chap?lang=eng&lettr=chapsect13&info0=13#zz13).

Questions Posted on the Discussion Board

B3.1 The current Online Standards do not contain any standards or guidelines pertaining to literacy. The literature review commissioned by PORD recommends the following:

> When it comes to literacy, questionnaires should use plain language to the extent possible and warranted.

Should there be a literacy standard or guideline in the Online Standards? If so, what do you suggest it be, and is it a standard, or a guideline (literature review recommendation is a guideline). And, should there be examples of "plain language", or not? Or, a reference to The Canadian Style Guide on *Plain Language*?

B4. Accessibility and Telephone Surveys

The current Telephone Standards do not contain any standards or guidelines pertaining to the accessibility of telephone surveys to people with disabilities or to other people who may have difficulty completing a telephone survey.

Questions Posted on the Discussion Board

B4.1 There are currently no accessibility standards in the Telephone Standards. Should the updated Telephone Standards include any accessibility standards or guidelines? If so, what should these be? For example, should there be options for TTY or alternative modes such as online or mail? An option for proxy respondents?

B4.2 Should the Telephone Standards have some sort of "plain language" standard or guideline, analogous to a literacy standard/guideline discussed by the Panel in Question B3? If so, what do you suggest it should be?

C. Surveys and Social Media

For this section, please scan MRIA's *Appendix C – Guideline on Social Media Research*. Note that the Panel focus is limited to usage of social media only in connection with conducting online or telephone surveys, and that also meet PORD's definition of public opinion research[54]. Under this

[54] PORD's definition of public opinion research can be found at: https://www.canada.ca/en/treasury-board-secretariat/services/government-communications/public-opinion-research-government.html

definition (1) there must be attitudinal/opinion questions in the research, and (2) the research must be based on asking questions. So, research consisting of web scraping is not considered to be public opinion research.

The focus of the Advisory Panel is on revising and updating the standards for telephone surveys and online surveys.

The *Introduction* to the MRIA's *Appendix C – Guideline on Social Media Research* – states:

> The concept of consumers generating their own content on the internet has become ubiquitous. This has created new opportunities for researchers to observe, interact and gather information. Many techniques have been developed to leverage social media such as community panels, crowd-sourcing, co-creation, netnography, blog mining and web scraping. It is likely that many more will evolve over the coming years as the Internet continues to change.

Many of the research possibilities referred to above fall outside the scope of PORD's Online and Telephone Standards, because the activities do not qualify as public opinion research surveys. However, particularly in the case of market research online communities (MROCs), it is possible to do an online survey and perhaps even a telephone survey. And, it is possible that other types of social media venues can be platforms for sampling and administering surveys.

If an online or telephone survey is done using a social media venue as the sample source, and perhaps additionally as the medium for administering an online survey, the research project would have to conform to all of the relevant Standards, that is, the Online Survey Standards or the Telephone Survey Standards.

The question is whether anything needs to be added to the Standards to cover online or telephone surveys that make use of social media – meaning, the sample is sourced from a social media, and the additional possibility that the survey is administered via the social media venue.

One perspective is that nothing really needs to be added to the Standards for social media-based surveys. The Standards lay out requirements for proposal documentation, questionnaire design, sampling, retaining public confidence, data collection, data security, etc. Adhering to these standards could be considered sufficient for an acceptable social media-based survey conducted for the Government of Canada.

A different perspective is that there are arguably some issues specific to social media-based surveys that are not clearly covered by the Standards. The MRIA's Code of Conduct includes *Appendix C – Guideline on Social Media Research*, which is based on ESOMAR's guideline. This code of conduct is intended to cover social media research broadly, not just surveys, but surveys are within its scope (MROCs are of particular interest in this context).

The MRIA guideline describes "key principles" for researchers. These principles are consistent with the Standards, but are explained specifically in terms of application to social media:

2.1 Distinguishing market research as the purpose
2.2 Conforming to the law

2.3 Consent and notification
2.4 Protecting identifiable data
2.5 Ensuring no harm
2.6 Children, young people and persons with functional cognitive disability
2.7 Reputation of the industry
2.8 Reporting

Also of particular interest for the survey Standards is Section 3, *Recommendations for Specific Social Media*, and within this, Section 3.2 *Private social media area issues* and Section 3.3 *Market research social media area issues.* These address aspects of permission, informed consent and privacy specific to these types of social media venues used for research.

Questions for the Panel

C1. It would be possible to add to the Standards a requirement to adhere to the MRIA's *Appendix C – Guideline on Social Media Research* in cases where a survey sample is sourced from a social media venue, and perhaps additionally fielded on a social media venue.

For example, in Section 5 *Retaining Public Confidence*, an item could be added to Section 5.1 *Respondent Rights*, such as:

> 5.1.6. Surveys using social media as the sample source or medium of survey administration
> 1) A survey that uses a social media venue as a sample source or to administer the survey must conform to all of the standards in this document, and to Appendix C *Guideline on Social Media Research* of the MRIA's *Code of Conduct for Market and Social Research.*

Do you think it is necessary or useful to add something like the above to the Standards, or not? If yes, do you have any suggestions for revisions to the above proposed wording?

C2. The Government of Canada holds itself to high standards for the conduct of public opinion research. In this context, adherence to the MRIA *Guideline on Social Media Research* would be a minimum requirement. Is the MRIA social media guideline sufficient as it pertains to online or telephone public opinion research surveys, or is there some part of it that you think should be made stronger for Government of Canada surveys that make use of a social media venue?