Nonparticipation (Unit Nonresponse) In Surveys:
A Practitioner's Guide to the Conceptualization, Impact of, and Adjustment for Unit Nonresponse

Pamela Woitschach, Ph.D.

Post-Doctoral Fellow, BC Office of Patient-Centred Measurement

BC SUPPORT Unit

Post-Doctoral Research Fellow, UBC-Paragon Research Initiative

University of British Columbia

Bruno D. Zumbo, Ph.D.

Professor & Distinguished University Scholar

Tier 1, Canada Research Chair in Psychometrics and Measurement; &

Paragon UBC Professor of Psychometrics and Measurement

University of British Columbia

Financial Support from:

The UBC-Paragon Research Initiative and, Professor Bruno D. Zumbo, Paragon UBC Professor of Psychometrics and Measurement

How to Cite This Report

Woitschach, P., & Zumbo, B.D. (2021, March 31). *Nonparticipation (Unit Nonresponse) In Surveys: A Practitioner's Guide to the Conceptualization, Impact of, and Adjustment for Unit Nonresponse [webinar]*. Vancouver, B.C.: British Columbia Patient-Centred Measurement.

Purpose of the Webinar

- This webinar is about conducting data analyses of survey data when some data values are missing- that is, some people do not respond to your survey. Later you will learn that this is referred to as 'unit nonresponse.'
- We will discuss how …
  - Missing data may complicate the interpretation and analysis of survey data
  - Unit nonresponse may alter/distort your conclusions from your analyses, and by how much
  - With the use of guiding principles and a framework for data analysis, there are workable statistical analyses that can go a long way toward adjusting for unit nonresponse.

Intended Audience

- Data analysts and health researchers who analyze survey data.
- We anticipate that the webinar audience has completed course work in health data analysis, or social and psychological data analysis covering the following topics:
  - Introduction to quantitative methods (e.g., different kinds of research designs, internal validity)
  - Basic probability theory in a health of social science course
  - Descriptive statistics (mean, median, mode)
  - Inferential statistics such involving p-values, confidence intervals
  - Correlation, basic linear and logistic regression with one or more predictors.
- We assume minimal background knowledge about survey methodology.
  - We will be motivating and describing the survey method with little to no mathematical equations.
- Statistical software for data analysis.
  - Able to install R and the R Studio (https://cran.r-project.org/mirrors.html)
  - Basic knowledge of R and R Studio (https://www.r-project.org/)

Learning Objectives

- Research **design and sampling** so that the audience will be able to situate the research problem into a larger context of research methods.
- Key concepts needed to understand unit nonresponse, including how it is different than other **types of missing data**.
  - Terms like missing completely at random (**MCAR**), **MAR**, **MNAR**, and what is meant by 'missing data is nonignorable'.
- The **impact of unit nonresponse** on typical data analyses; answering questions such as when is unit nonresponse an issue and how much is the practical impact.
  - We will demonstrate how much the results of statistical analysis of complete data (only those who responded) can alter your conclusions and when.
- **Demonstrate a practical approach to adjust for unit nonresponse**.
  - The key idea is the probability of a sampled respondent replying to the survey;
  - We do not have the probability of a sampled respondent replying to the survey; therefore, we need to estimate using any variables we may have for both respondents and non-respondents.

Structure / Outline

The Unit Nonresponse webinar is organized in three parts that will allow our audience to be engaged with the information they consider to be relevant to their interests.

It is important to keep in mind that the webinar was created for *data analysts and applied researchers* who work with data to respond to research questions but who are not steeped in the research literature in statistical science and mathematical statistics.

The webinar topics are as following:

Part I: Is focused on introducing the concepts leading to an understanding of survey research and unit nonresponse.

Part II: Demonstrates the impact of ignoring unit nonresponse in survey research and using complete-case analysis.

Part III: Aims to present a sample of the tools often used to adjust for unit nonresponse. Moreover, Part III will include a demonstration of the Propensity Score Weighting approach on simulated data.

Nonparticipation (Unit Nonresponse) In Surveys:
A Practitioner's Guide to the Conceptualization, Impact of, and Adjustment for Unit Nonresponse
Webinar Part 1
Introduction

Pamela Woitschach, Ph.D.

Post-Doctoral Research Fellow, BC Office of Patient-Centred Measurement

BC SUPPORT Unit

Post-Doctoral Research Fellow, UBC-Paragon Research Initiative

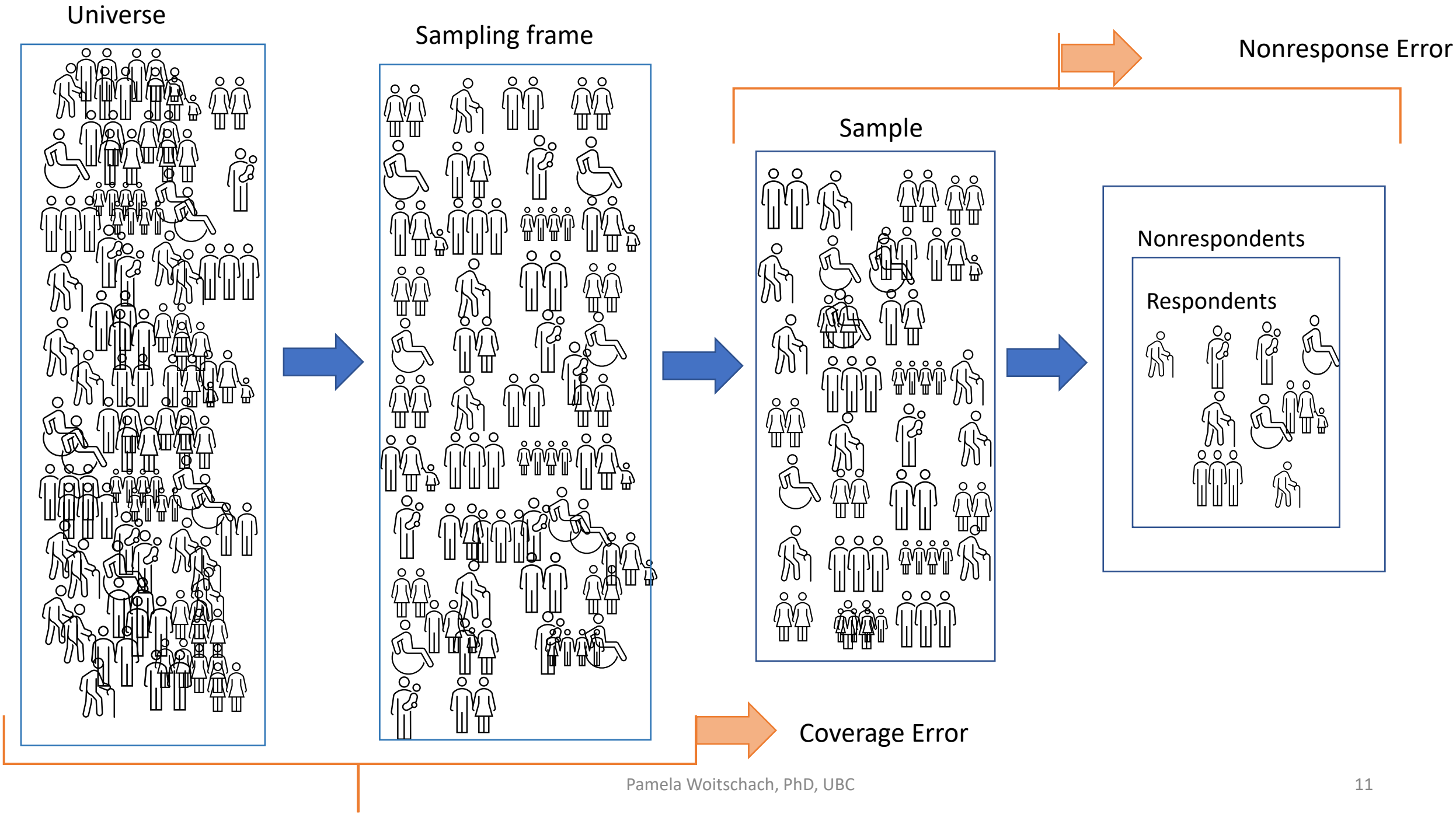
University of British Columbia

Outline of Part 1

- Motivation of research
- Research design & sampling
  I. Experimental, Quasi-experimental & Observational studies: Introduction to surveys
  II. Sampling and Survey design
  III. Separation between measure and survey
  IV. What do they do with those measurements?
- Setting the stage: definitions, and important concepts
  I. Missing data patterns and mechanisms
  II. Unit not-response definition. Is unit nonresponse ignorable?
- Transitioning to Part 2 of the Webinar

Motivation of Research (1/5)

- In British Columbia, the measurement of patient-reported experiences (PREMs) and outcomes (PROMs) have been collected for the last 16 years to understand the experiences of people who use BC's healthcare services, to enhance public accountability, and to support quality improvement.
- A detailed methodological design of each survey had been followed over the years. However, the participation observed across and, mainly, in the last two surveys (The Emergency Department Sector Survey-2018 & the Acute Inpatient Sector Survey-2016/17) ranged from 35% to 46%, leaving an average of 50% of sampled people who for several reasons decided not to participate.


Universe
Sampling frame
Nonresponse Error
Sample
Nonrespondents
Respondents
Coverage Error

Motivation of research (2/5)

- Nonrespondents' groups are often conformed by people who are unable to contact (*noncontact*) or unwilling to participate (*refusal*). Nonresponse can also be due to *accidental loss of the data*, or *hard to reach participants* due to their sociodemographic situation. Nonrespondents can also be the people who prematurely terminate an interview.

- With the widespread use of data as a tool for program evaluation and decision-making, the channels in where surveys are delivered have changed from face to face, letters, fax, and telephone to online, computerized, and cell phone-based surveys. Making the data collection more efficient and accessible.

Motivation of research (3/5)

- Nonresponse can also be classified depending on the relationship between the *response mechanism* (MCAR, MAR, and MNAR) and the variable of interest resulting in *ignorable* and *nonignorable* nonresponse.
- *Nonresponse bias* is a multiplicative function between the *response rate* and the difference between survey respondents and nonrespondents with respect to the variable of interest (Grooves, 2006).
- It is not always feasible to know the *nonrespondent's characteristics* to infer differences on the variable of interest between both groups. Thus, the *response rate,* is one of the most frequent indicators used for survey quality as it is easy to calculate, and it does not require information from the nonrespondent's group.

Motivation of research (4/5)

- As a data analyst, we aim to infer characteristics of our population (the data we wish to have) from the sample's responses. However, far too often, we encounter ourselves with a final dataset that only contains a group of respondents (the *data we have/observed data*) and several nonrespondents (from whom we do not often know more than a few sociodemographic variables).

- Most of the statistical software is built to work with complete case analyses, so we tend to work only with the observed data. Still, we do not know how similar our observed data is to the data we do not have.

Motivation of research (5/5)

<u>The study of nonresponse can be divided in three main areas:</u>

(1) Documenting how often and why nonresponse occurs and how to increase response rates (Sociologist perspective).

(2) How to address the presence of nonresponse (Statistical perspective).

(3) Documenting the impact of unit nonresponse in survey research (Methodological perspective).

One of the difficulties preventing a deeper understanding of nonresponse in surveys is the complexity of the survey process

(Brick, 2013. p.347).

Research design & sampling

Build a foundation on survey research to differentiate it from other types of research designs. Observe their characteristics, type of instruments applied, and the claims that can be made from the survey results obtained.

Experimental, Quasi-experimental & Observational studies: Introduction to surveys

The use of randomization in samplings. As we are interested in characteristics of a certain population, and this group is either impossible or at least difficult to reach, we tend to estimate their characteristics based on a sample that was randomly selected from the population. And from which we assumed the population is represented.

It's a matter of control… or how do we apply this control. The selection of participant of observations could be controlled by social characteristics.

"Experiments are strong on control through randomization; but they are weak on representation. Surveys are strong on representation, but they are often weak on control" Kish (1959, p.332)

TABLE 1.1 The Vocabulary of Experiments

Experiment: A study in which an intervention is deliberately introduced to observe its effects.
Randomized Experiment: An experiment in which units are assigned to receive the treatment or an alternative condition by a random process such as the toss of a coin or a table of random numbers.
Quasi-Experiment: An experiment in which units are not assigned to conditions randomly.
Natural Experiment: Not really an experiment because the cause usually cannot be manipulated; a study that contrasts a naturally occurring event such as an earthquake with a comparison condition.
Correlational Study: Usually synonymous with nonexperimental or observational study; a study that simply observes the size and direction of a relationship among variables.

Note: Reprinted from *Experimental and quasi-experimental designs for generalized causal inference (p.13),* by Shadish, Cook, & Campbell, 2002, Houghton Mifflin.

What is survey research?

"Survey research is a type of research design indented to obtain information through the application of a questionnaire to participants that conform a sample that is drawn from a population framework where the probability of selection is known "

"A survey is any activity that collects information in an organised methodological manner about characteristics of interest from some or all units of a population using well-defined concepts, methods and procedures, and compiles such information into a useful summary form".

Statistics Canada. *Survey Methods and Practices* (2003, p.1).

Survey research and sampling

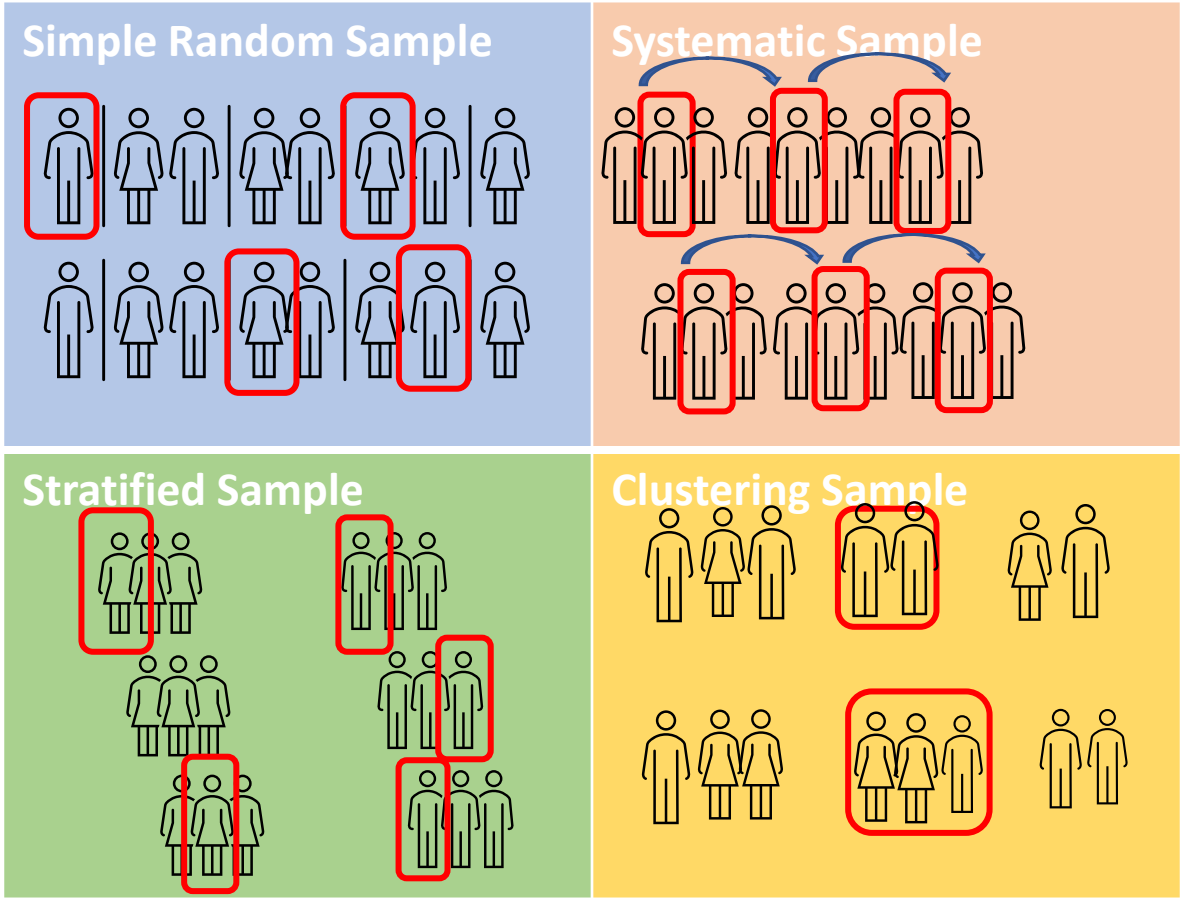
- Frame lists, selection, recruitment.
- Census and sample.
- Classification of Sample Designs.
- Ronald Fisher's (1922) Model-based Framework for Inference.
- Jerzy Neyman's (1934) Design-based Framework for Inference.

Sampling Plan	Method of Inference	
	Design-based	Model-based
Probability Sample	A	B
Model-dependent Sample	C	D

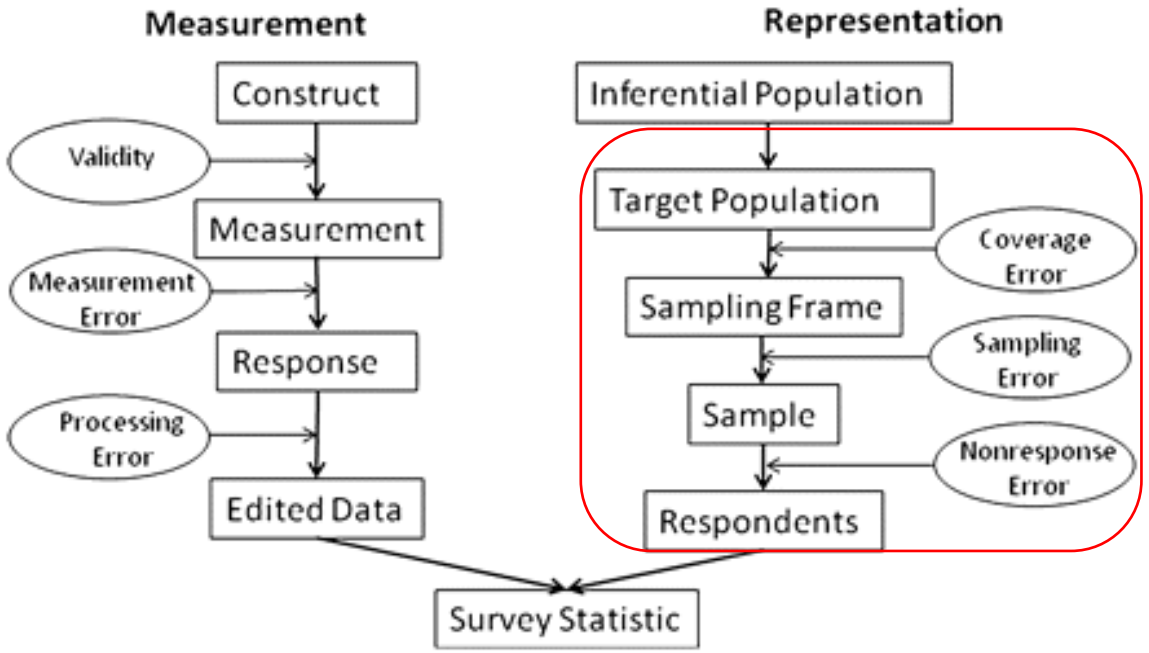
Note: Extracted from Heeringa, S., West, B., Berglund, P. (2010) Applied Survey Data Analysis.

Survey research and sampling

- Sampling techniques (Simple random, systematic sampling, stratified sampling, clustering sampling).
- Deficiencies in a sample (nonresponse, missing data, noncoverage).
- Compensation weighting.



Total survey error components



Note: Extracted from Total Survey Error: Past, Present, and Future, Groves, & Lyberg, 2010, *Public Opinion Quarterly, 74* (5), pp.849–879, https://doi.org/10.1093/poq/nfq065

Deming, 1944
Nonresponse

Groves, 1989
Attention to the dinstinction between errors of nonobservation and errors of observation.

Biemer & Lyberg, 2003.Nonresponse

Cochran, 1953
Nonresponse errors

Lellser & Kalsbeek, 1992
Nonresponse errors

Groves et al., 2004
Nonresponse error as a part of total survey error

Drawing conclusions from the sample

- Defining the population.
- Frame identification does not contain all the population.
- Sample selection.
- Estimation.
- Standard error.
- Confidence interval.

Measures of data quality

- Response rate: is one of the most frequently used indicators of survey quality (Koch & Blohm, 2016). The response rate is the ratio of the number of respondents to the number of eligible units (Brick & Montaquilla, 2009).

- In linear statistics Nonresponse bias is a multiplicative function between nonresponse rate and the difference between respondents and nonrespondents with respect to a certain survey variable (Koch & Blohm, 2016).

Separation between measure and survey

- Groves (1989) attempted to link error notions from the TSE formulation with psychometric and econometrics. Some concepts from TSE were missed in the psychometric literature as well as concepts from TSE were also avoided in psychometric literature (Groves, 2010).
- Reporting errors in surveys arise from problems in the underlying cognitive processes through which respondents generate their answers to survey questions (Tourangeau, 2003).
- Based on Zumbo's (2007) ecological model for responding to survey, the statistical model is shaped by the ecological framework, or more generally, a validation program of research is applied to explain the variation in responding to survey questions and questionnaires. The ecological methods work to establish and support the explanation for item responding.

What do researchers do with those measurements?

(1) Monitoring.
(2) Psychometrics of the measure.
(3) Compare and classify relevant sub-groups.
(4) Describe and correlate variables among groups.
(5) Build and apply statistical models.

Statistical model aims to draw inferences from calibration samples to the respective populations from which these were drawn (Zumbo, 2007, p. 61).

Setting the stage: definitions, and important concepts

To set up a clear understanding of the concepts related to missing data patterns and mechanisms, nonresponse, and the relevance of model specification on survey research designs that will allows us to understand the next sections of the webinar.

Missing data occur in survey research because an element in the target population is not included on the survey's sampling frame (noncoverage), because a sampled element does not participate in the survey (nonresponse) and because a responding sampled element fails to provide acceptable responses to one or more of the survey items (item nonresponse).

Defining Unit non-response (1)

"Total nonresponse occurs when no survey data are collected for an element selected for the sample. The nonresponse results from refusals to participate in the survey, noncontacts (not at homes), and other reasons such as a language or being too ill to participate. Compensation for total nonresponse is usually made by means of weighting adjustments in which respondents are assigned greater weight in the analysis in order to represent nonrespondents"

(Brick & Kalton, 1996, p1).

Defining Unit non-response (2)

"There are two types of nonresponse: item (or partial) nonresponse and total nonresponse. Item nonresponse occurs when information is provided for only some items, such as when the respondent responds to only part of the questionnaire. Total nonresponse occurs when all or almost all data for a sampling unit are missing".

Statistics Canada.

Survey methods and practices (2003, p.34).

Defining Unit non-response (3)

"Most surveys suffer from nonresponse. This is the phenomenon that sample elements do not provide the required information. Non-response may seriously affect the quality of the outcomes of a survey. Estimates of population characteristics will be biased if, due to non-response, some groups in the population are over-or underrepresented, and these groups behave differently with respect to the survey variables".

Bethlehem et al. (2008, p. 1)

Defining Unit non-response (4)

"Non-participation is a common phenomenon which, when differences between participants and non-participants are considerable, may reduce the external validity of the survey and increase risk of drawing false conclusions about the health status of the population".

Carlsson et al. (2006, p. 132)

Representativity of a postal public health questionnaire survey in Sweden, with special reference to ethnic differences in participation

Missing survey data

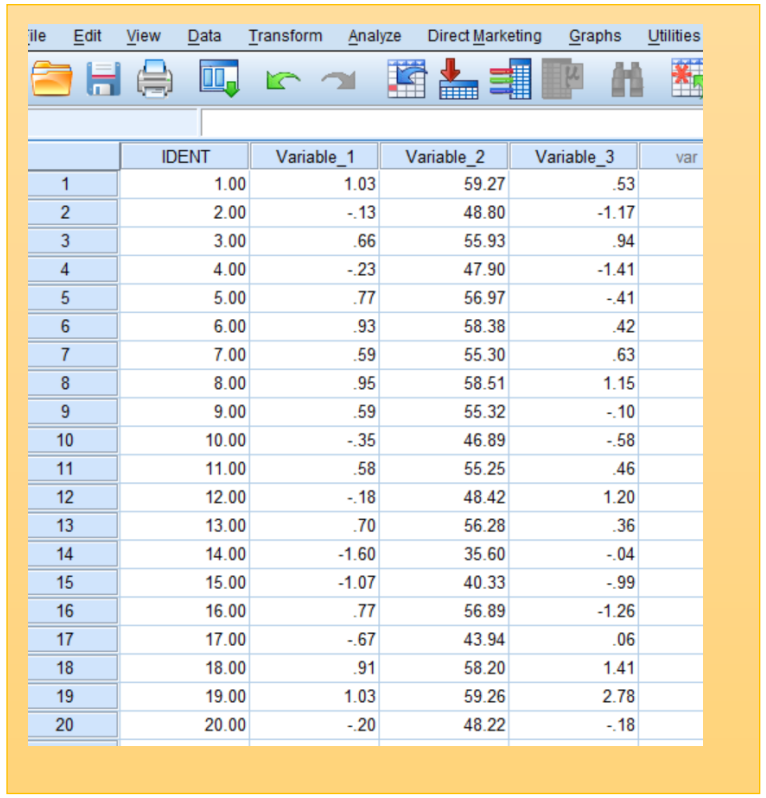
"There are two types of nonresponse: item (or partial) nonresponse and total nonresponse. Item nonresponse occurs when information is provided for only some items, such as when the respondent responds to only part of the questionnaire. Total nonresponse occurs when all or almost all data for a sampling unit are missing".

Statistics Canada. Survey methods and practices (2003, p.34).

Type of missing survey data	Definition	Causes	Type of compensation
Total nonresponse	No survey data are collected for an element selected for the sample	Results from refusals to participate in the survey, noncontacts (not at homes), and other reasons such as a language barrier, deafness or being too ill to participate.	Weighting
Noncoverage	Some elements in the population of inference for the survey are not included in the survey's sampling frame		Weighting
Item nonresponse	When a sampled element participates in the survey but fails to provide acceptable responses to one or more of the survey items	A respondent refuses to answer an item on the grounds that it is too sensitive, does not know the answer to the item, gives an answer that is inconsistent with answers to other items and hence is deleted in editing, or because the interviewer fails to ask the question or record the answer	Imputation
Partial nonresponse	Involves a substantial number of item nonresponses	When a respondent cuts off the interview in the middle, when a respondent in a panel survey fails to provide data for one or more of the waves of the panel, or when a respondent t in a multiphase survey provides data for some but not all phases of data collection	Either weighting or imputation

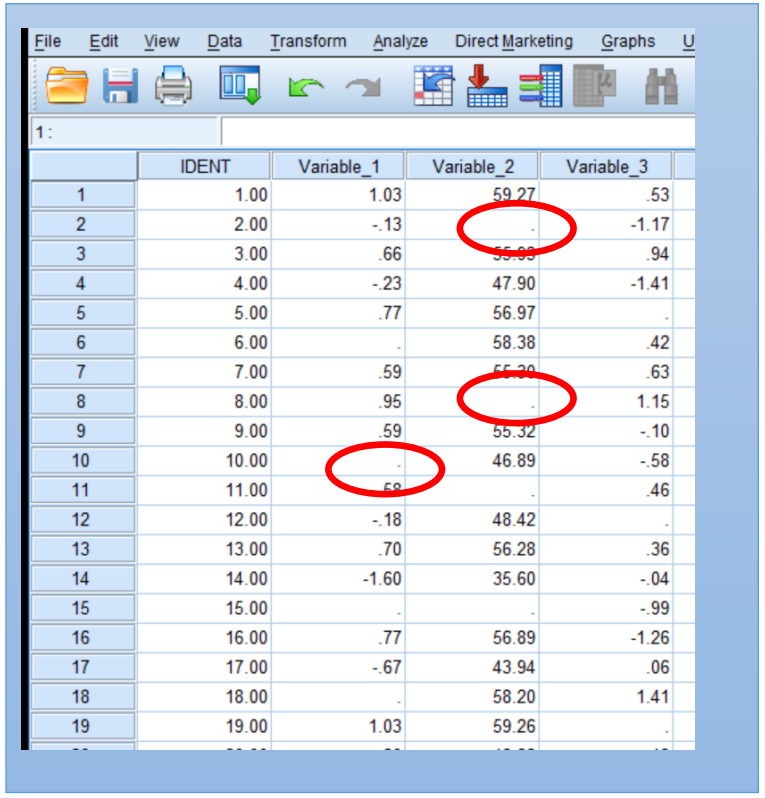
Note. **Adapted from Brick & Kalton, 1996**

Ideal case- no missing data



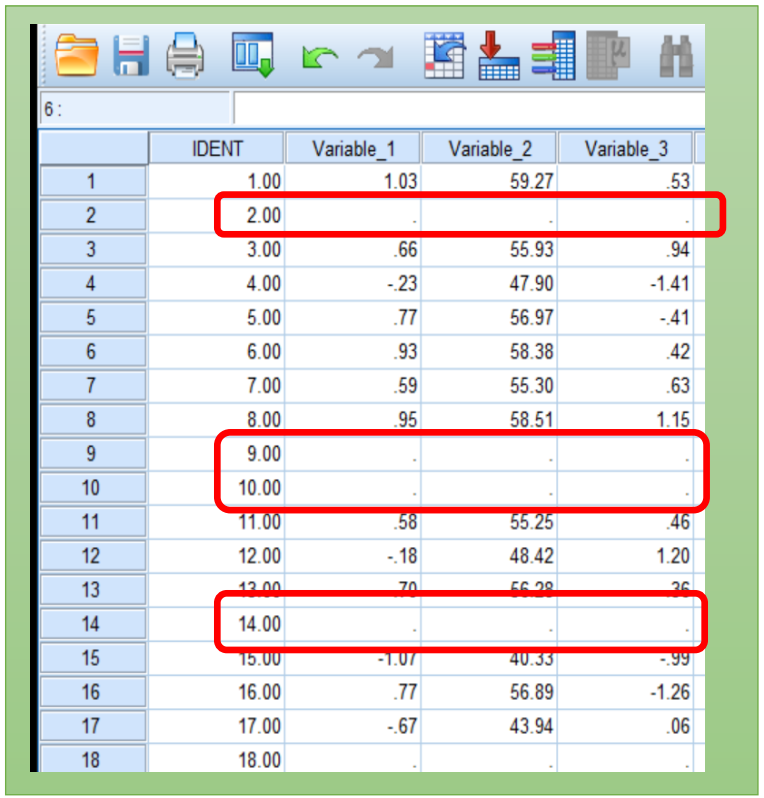
	IDENT	Variable_1	Variable_2	Variable_3	var
1	1.00	1.03	59.27	.53	
2	2.00	-.13	48.80	-1.17	
3	3.00	.66	55.93	.94	
4	4.00	-.23	47.90	-1.41	
5	5.00	.77	56.97	-.41	
6	6.00	.93	58.38	.42	
7	7.00	.59	55.30	.63	
8	8.00	.95	58.51	1.15	
9	9.00	.59	55.32	-.10	
10	10.00	-.35	46.89	-.58	
11	11.00	.58	55.25	.46	
12	12.00	-.18	48.42	1.20	
13	13.00	.70	56.28	.36	
14	14.00	-1.60	35.60	-.04	
15	15.00	-1.07	40.33	-.99	
16	16.00	.77	56.89	-1.26	
17	17.00	-.67	43.94	.06	
18	18.00	.91	58.20	1.41	
19	19.00	1.03	59.26	2.78	
20	20.00	-.20	48.22	-.18	

Item nonresponse



	IDENT	Variable_1	Variable_2	Variable_3
1	1.00	1.03	59.27	.53
2	2.00	-.13	.	-1.17
3	3.00	.66	55.93	.94
4	4.00	-.23	47.90	-1.41
5	5.00	.77	56.97	.
6	6.00	.	58.38	.42
7	7.00	.59	55.30	.63
8	8.00	.95	.	1.15
9	9.00	.59	55.32	-.10
10	10.00	.	46.89	-.58
11	11.00	.58	.	.46
12	12.00	-.18	48.42	.
13	13.00	.70	56.28	.36
14	14.00	-1.60	35.60	-.04
15	15.00	.	.	-.99
16	16.00	.77	56.89	-1.26
17	17.00	-.67	43.94	.06
18	18.00	.	58.20	1.41
19	19.00	1.03	59.26	.

Unit nonresponse



	IDENT	Variable_1	Variable_2	Variable_3
1	1.00	1.03	59.27	.53
2	2.00	.	.	.
3	3.00	.66	55.93	.94
4	4.00	-.23	47.90	-1.41
5	5.00	.77	56.97	-.41
6	6.00	.93	58.38	.42
7	7.00	.59	55.30	.63
8	8.00	.95	58.51	1.15
9	9.00	.	.	.
10	10.00	.	.	.
11	11.00	.58	55.25	.46
12	12.00	-.18	48.42	1.20
13	13.00	.70	56.28	.36
14	14.00	.	.	.
15	15.00	-1.07	40.33	-.99
16	16.00	.77	56.89	-1.26
17	17.00	-.67	43.94	.06
18	18.00	.	.	.

Complete-case Analysis (CCA)

When a researcher is estimating a model (e.g., linear regression), most of the statistical software packages use listwise deletion by default.

In CCA the analysis restricted to individuals with complete information on all variables of the main analysis. When we use CCA we automatically obtained a less efficient statistical result.

- MCAR assumption.
- Adequate amount of data available for analysis.
- Standard error is larger.

However, in a systematic review of 262 studies published in 2010 in three leading epidemiologic journals. Complete-case analysis was reported in 81% of the studies (Eekhout, et al., 2012, p.729).

Why people might respond or not to a survey

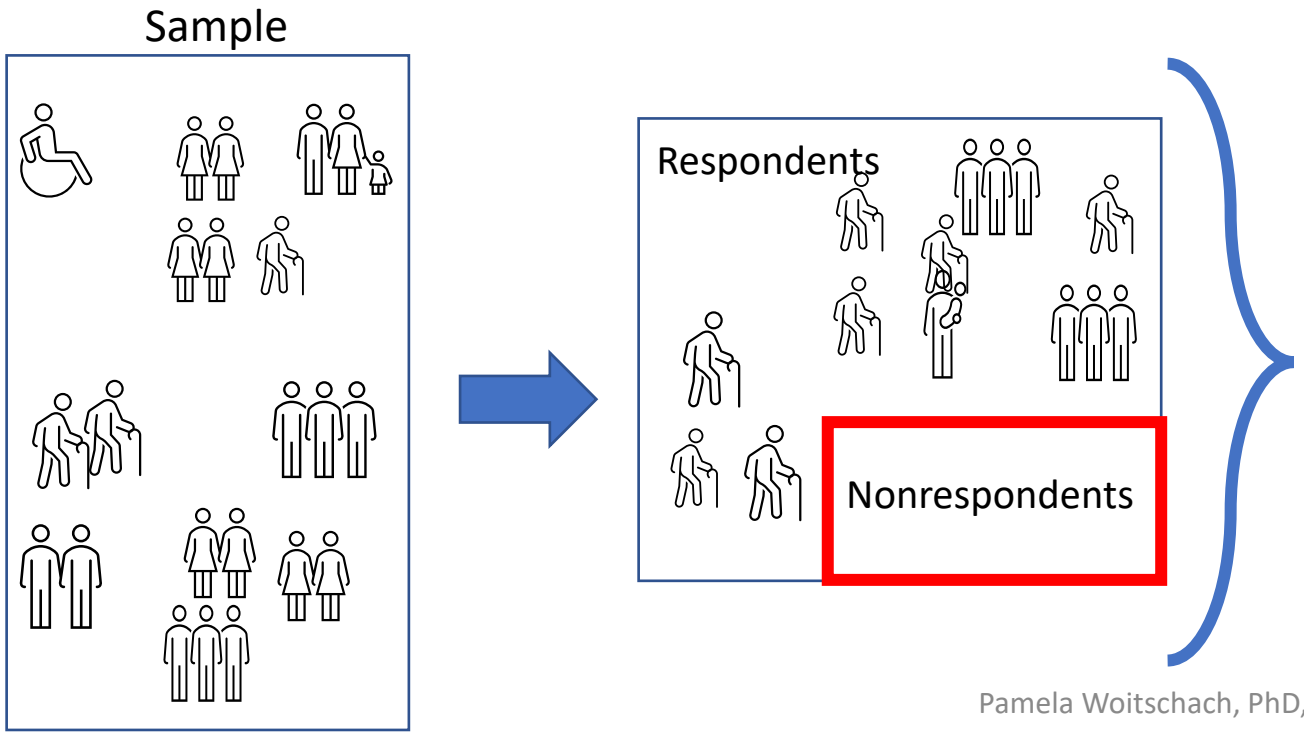
"People respond to a survey when they conclude that the rewards outweigh the costs. Societal factors also enter in" National Research Council.

Nonresponse in Social Science Surveys: A Research Agenda. (2013, p32)

Accessibility

Amenability

Administrative issues



In the presence of nonresponse, the set of respondents can be viewed as a second-phase sample, the only difference is that the probabilities are unknown (Haziza, 2009).

The causes for nonresponse can be numerous, and it can vary between surveys, and within participants. So, it is practically unrealistic to describe all of the possible reasons (Haziza, 2009).

Missing data patterns and mechanisms (MCAR, MAR, MNAR)

- Missing data patterns, describes which values are observed in the data matrix and which values are missing, and the missing data mechanism, concerns to the relationship between missingness and the values of variables in the data matrix (Little & Rubin, 2002).
- Whether a unit responds or not is treated as a random event when response is categorized as one of these.
- Responding could be considered deterministic, but random or stochastic response is the formulation behind the nonresponse adjustments used.

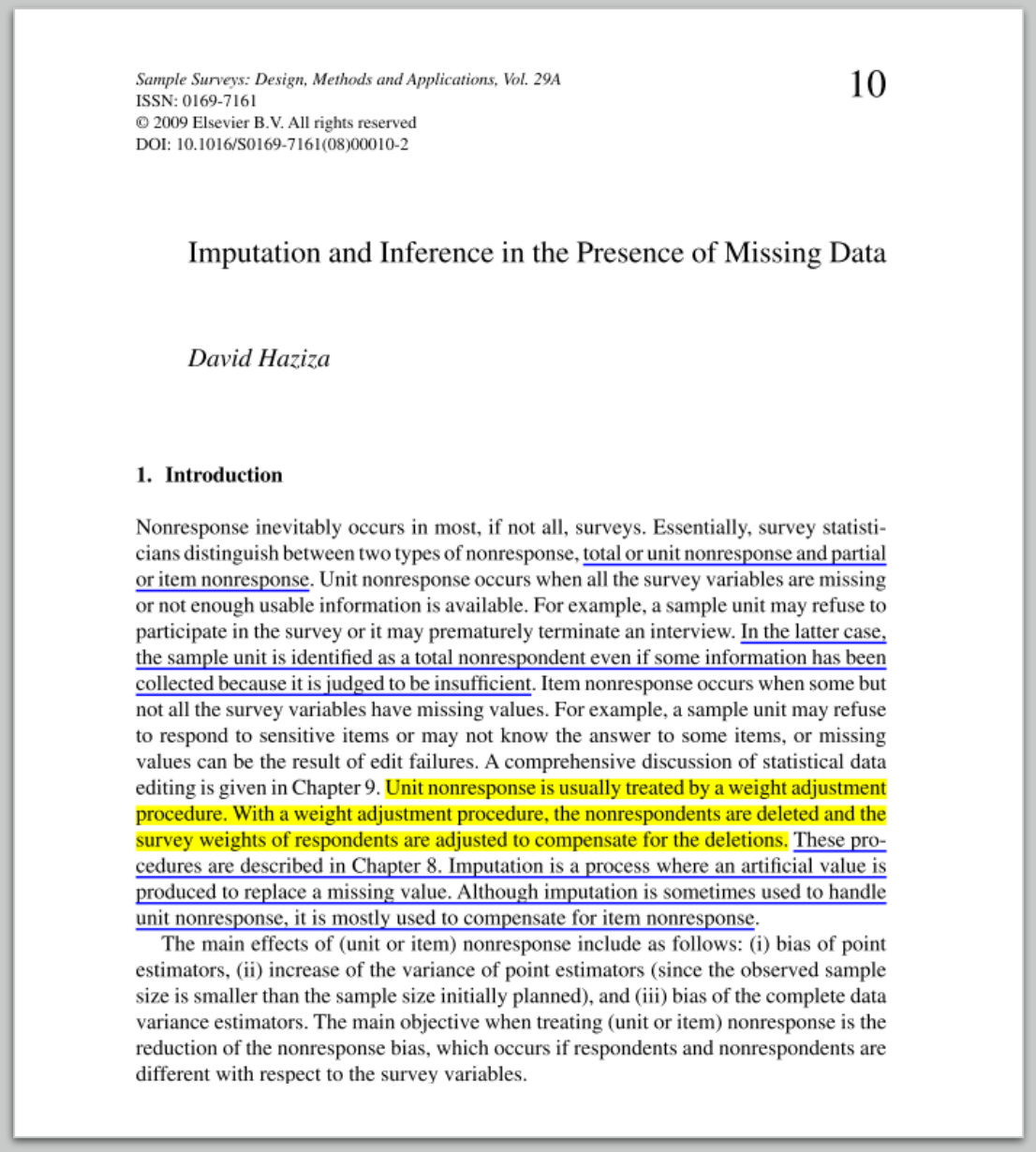
Missing data mechanism		
Missing completely at random (MCAR)	If missingness does not depend on the values of the data. Note that this assumption does not meant that the pattern itself is random, but rather that missingness does not depend on the data values.	Every unit has same probability of responding. Responding is just an extra stage of Bernoulli sampling. No weight adjustment needed for means; one overall adjustment need for totals.
Missing at random (MAR)	When missingness depends only on the components that are observed, and not on the components that are missing.	Probability of responding depends on covariates. Adjustment possible if covariates known for both respondents and nonrespondents.
Missing not at random (MNAR)	Is when missingness depends on the missing values in the data matrix.	Also know as non ignorable nonresponse. Probability of responding depends on analytic variables and possibly covariates. Adjustment difficult or impossible.

When is unit nonresponse ignorable?

- Ignorable = MCAR, and MAR.

- Non-ignorable=MNAR.

In practice, the ignorability of the nonresponse mechanism is assumed because it is generally impossible to test whether we are in presence of ignorable or nonignorable response except in the context of planned nonresponse (Haziza, 2009 p. 222).



Sample Surveys: Design, Methods and Applications, Vol. 29A
ISSN: 0169-7161

DOI: 10.1016/S0169-7161(08)00010-2

Imputation and Inference in the Presence of Missing Data

David Haziza

1. Introduction

Nonresponse inevitably occurs in most, if not all, surveys. Essentially, survey statisticians distinguish between two types of nonresponse, total or unit nonresponse and partial or item nonresponse. Unit nonresponse occurs when all the survey variables are missing or not enough usable information is available. For example, a sample unit may refuse to participate in the survey or it may prematurely terminate an interview. In the latter case, the sample unit is identified as a total nonrespondent even if some information has been collected because it is judged to be insufficient. Item nonresponse occurs when some but not all the survey variables have missing values. For example, a sample unit may refuse to respond to sensitive items or may not know the answer to some items, or missing values can be the result of edit failures. A comprehensive discussion of statistical data editing is given in Chapter 9. Unit nonresponse is usually treated by a weight adjustment procedure. With a weight adjustment procedure, the nonrespondents are deleted and the survey weights of respondents are adjusted to compensate for the deletions. These procedures are described in Chapter 8. Imputation is a process where an artificial value is produced to replace a missing value. Although imputation is sometimes used to handle unit nonresponse, it is mostly used to compensate for item nonresponse.

The main effects of (unit or item) nonresponse include as follows: (i) bias of point estimators, (ii) increase of the variance of point estimators (since the observed sample size is smaller than the sample size initially planned), and (iii) bias of the complete data variance estimators. The main objective when treating (unit or item) nonresponse is the reduction of the nonresponse bias, which occurs if respondents and nonrespondents are different with respect to the survey variables.

Haziza, D. (2009). Imputation and inference in the presence of missing data. In D. Pfeffermann & C. R. Rao (Eds.), *Handbook of Statistics. Sample Surveys: Design, Methods and Applications*. Elsevier. https://doi.org/10.1016/S0169-7161(08)00010-2

When the nonresponse mechanism is ignorable?

When a survey response mechanism depends on a variable of interest measured within the same survey and observed for only part of the sample, the situation is one of *nonignorable nonresponse* (Beaumont, 2000, p.131).

If *nonresponse is ignorable*, i.e., there is no correlation between response behavior and the survey variables. In consequence, estimates will be unbiased

(Bethlehem et al., 2008, p.1).

Missing data is *nonignorable*, when *nonresponse* depends on characteristics of interest that are either observed only on the respondents or are completely unobserved, which leads to data that are missing not at random (Matei & Ranalli, 2015, p. 145) .

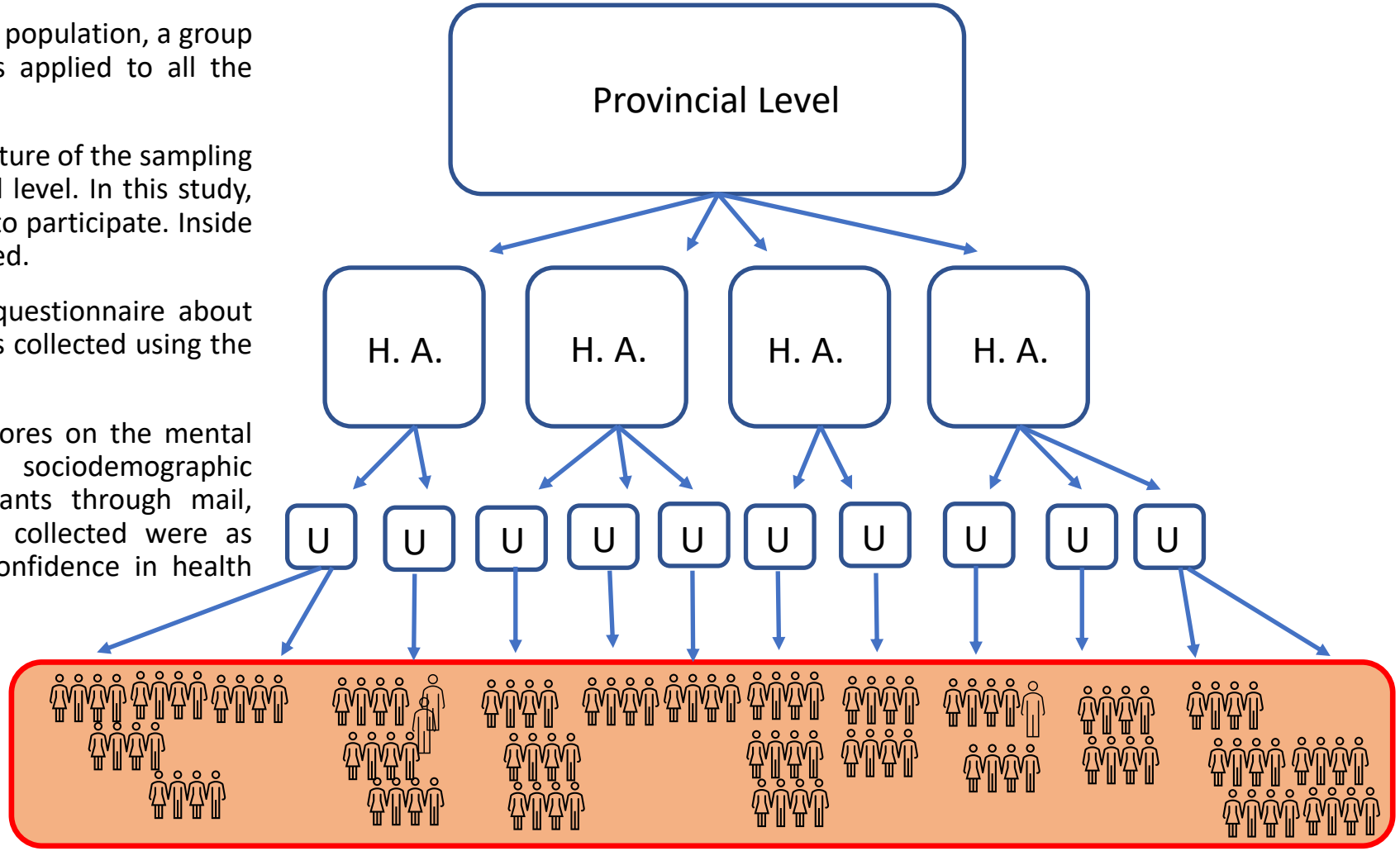
A Hypothetical Example to Conceptualize Surveys

To describe a specific mental health characteristic in a population, a group of researchers conducted a survey. The survey was applied to all the Health Authorities in the Province.

In this example, we will not consider the multilevel nature of the sampling design used, and we will only focus on the individual level. In this study, all the units on each Health Authority were selected to participate. Inside the units, a random sample of participants was selected.

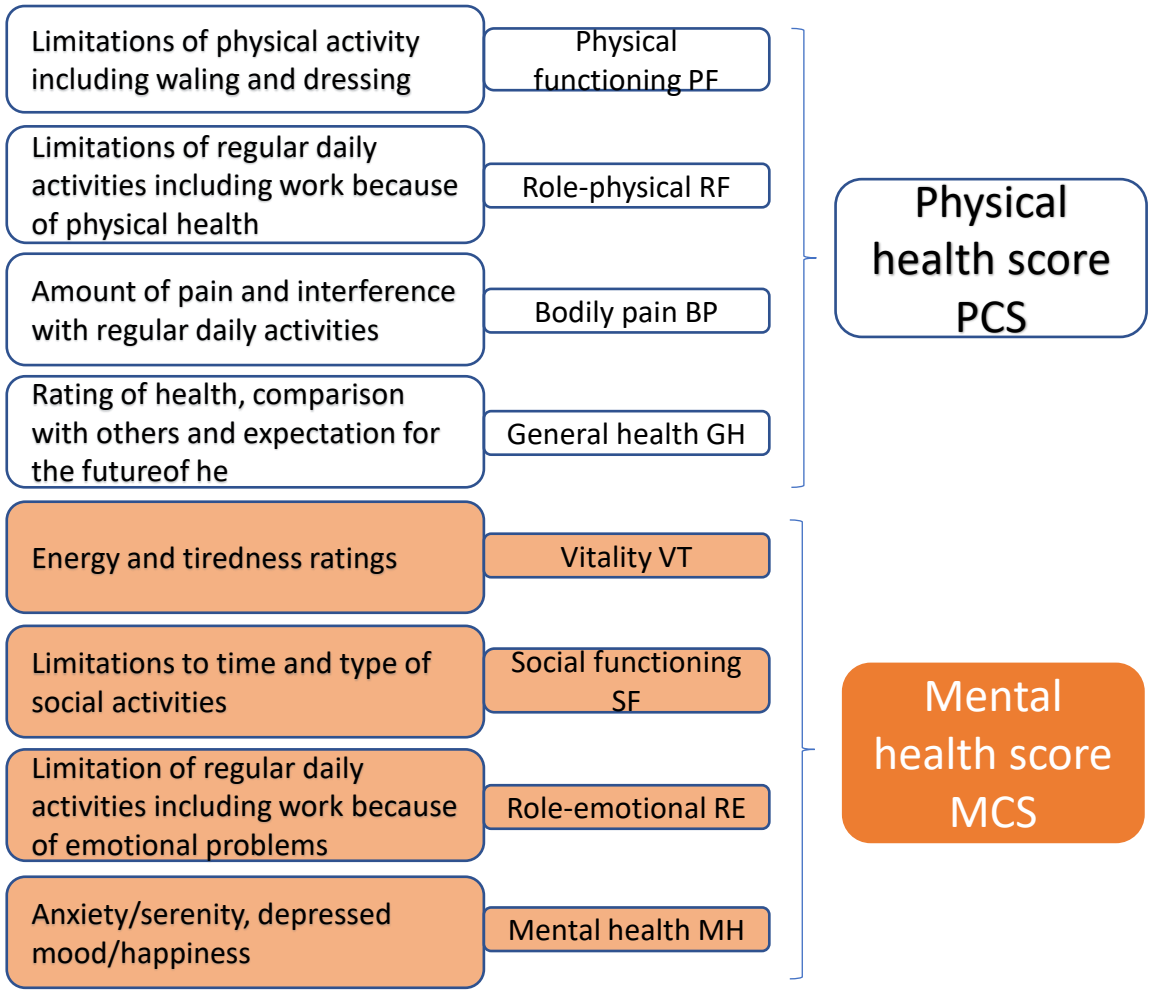
All participants were asked to respond to a short questionnaire about their sociodemographic. Their health information was collected using the SF-36 short form.

The study's focus is to evaluate the participants' scores on the mental health component summary (MCS) and their sociodemographic characteristics. The researchers contacted participants through mail, email, and telephone. Sociodemographic variables collected were as follows, age category, gender, area of residence, confidence in health providers, household size.



Our focus is at this individual level

Outcome measure (used in the simulations and demonstration in Parts 2 and 3)



SF-36 short form Health Survey

SF-36 aims to measure patient's health status and quality of life.

PCS: Composed by + weighting of 4 physical subscales (PF, RP, BP and GH), and – weighting of the psychological subscales (VT, SF, RE and MH).

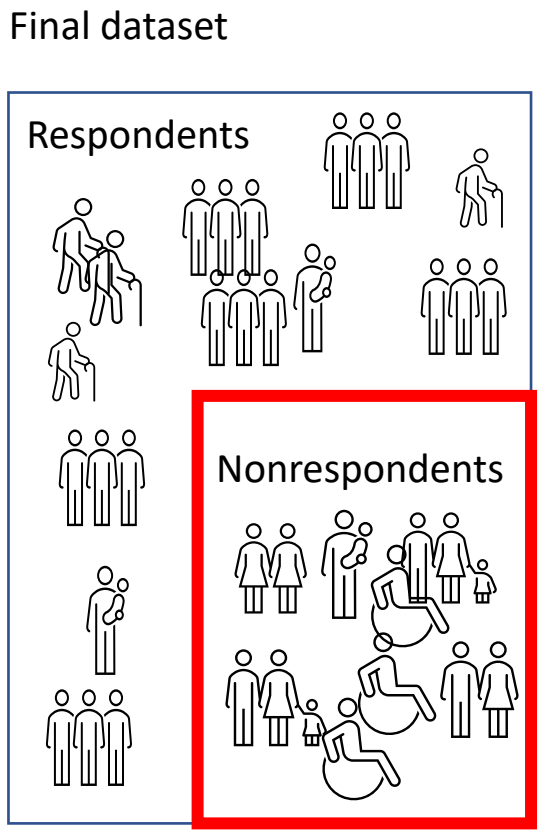
MCS: Is created by + weighting of the psychological subscales (VT, SF, RE and MH) and – weighting of 4 physical subscales (PF, RP, BP and GH).

A Hypothetical Example to Conceptualize Surveys

Now let's think about some of the concepts we have learned so far in this webinar.

- Population, Sampling frame, Sample, and final dataset (Respondents and Nonrespondents).

- Outcome measure (Mental Health Score), and covariates (age category, gender, area of residence, confidence in health providers, household size).



Data Linkage, also known as data matching, duplicate detection, or entity resolution, is the process of identifying and aggregating records from one or more datasets
(Chi, et al., 2016, p.199)

Dataset A: Containing nonresponse

8 : Age

	Unique_Identifier	MCS	Age
1	1.00	59.27	47
2	2.00	48.80	59
3	3.00	55.93	47
4	4.00	47.90	38
5	5.00	.	.
6	6.00	58.38	56
7	7.00	55.30	57
8	8.00	.	.
9	9.00	55.32	58
10	10.00	46.89	44
11	11.00	55.25	57
12	12.00	.	.
13	13.00	56.28	53
14	14.00	35.60	38
15	15.00	40.33	52
16	16.00	56.89	28
17	17.00	43.94	59
18	18.00	.	.
19	19.00	59.26	56
20	20.00	48.22	59
21	21.00	57.85	37

Dataset B: Containing ancillary data such as gender, age, are of residence, language spoken at home, etc.

Unique Identifiers

1 :

	Unique_Identifier	Sociodemographic_1	Sociodemographic_2	Sociodemographic
1	1.00	.53	-.30	47
2	2.00	-1.17	.99	58
3	3.00	.94	-.34	46
4	4.00	-1.41	-1.37	37
5	5.00	-.41	.82	57
6	6.00	.42	.71	56
7	7.00	.63	.73	56
8	8.00	1.15	-1.92	32
9	9.00	-.10	.94	58
10	10.00	-.58	-.65	44
11	11.00	.46	.83	57
12	12.00	1.20	.63	55
13	13.00	.36	.28	52
14	14.00	-.04	-1.34	37
15	15.00	-.99	.17	51
16	16.00	-1.26	-2.44	28
17	17.00	.06	.99	58
18	18.00	1.41	1.01	59
19	19.00	2.78	.63	55
20	20.00	-.18	1.00	58
21	21.00	1.08	-1.44	37

Data View | Variable View

When a dataset lacks of unique identifiers probabilistic data linkage may be used.

What do we do in the presence of unit nonresponse?

- Take a look at the problem: Why are they not responding?
- Is it because of the topic of the survey?
- Is it because of the time demands?
- Is it due to behavioral or sociodemographic factors?
- Is the unit nonresponse ignorable or nonignorable?
- Are there meaningful differences between the groups of respondents and nonrespondents?

Response Models

- There is a mechanism behind a nonresponse, a probability of a sample unit to respond to a survey. Furthermore, that probability is a latent variable that is therefore not directly observable and needs to be estimated.
- To estimate the probability of response of any given sampled unit, we use the ancillary data.
- Response models can be based on the information we have in our ancillary datasets (for respondents and nonrespondents), but they can also be based on a nonresponse theory.
- In this second case, Groves and Couper (1998) argued that theory-based response propensity models should begin by *discarding the notion that nonresponse is a fixed property*.

Putting all the pieces together

Putting all the pieces together (1/4)

- Up to this point, we reviewed some of the key points of a survey research process. We began by describing that every survey research's starting point is to specify a **population of interest** from where a **sample** will be selected. Various sampling designs and strategies are used to represent the population of interest and to achieve an optimal response rate.
- As researchers gathered data, they ended up with the final dataset that contains a proportion of **respondents, nonrespondents, and item missing data** among the respondents.

Putting all the pieces together (2/4)

- Nonrespondents' groups are often unable to contact or **unwilling** to participate. Nonresponse can also be due to **accidental loss of the data**, **inability to respond**, **or hard-to-reach participants** due to their sociodemographic situation.
- Nonresponse is commonly observed in surveys. However, when the differences between participants and nonparticipants are considerable, the survey's external validity can be reduced, and the risk of drawing a false conclusion about the population's characteristics can increase.
- When the nonresponse is based on the response pattern defined as **MNAR**, the nonresponse became **nonignorable**.

Putting all the pieces together (3/4)

- Moreover, it is essential to have a better understanding of why people decide not to participate. As we stated before, there is a **propensity of a sample unit responding to a survey**. That propensity is a latent variable that is therefore not directly observed and needs to be estimated.
- The response behavior is connected to the participant's interest and willingness to participate, and it is not a fixed property that can be generalized to all survey contexts. **Response models** can be based on ancillary information, but they can also be based on a nonresponse theory.

Putting all the pieces together (4/4)

- It is evident for us that nonresponse is a phenomenon often studied from a sociological and statistical perspective. Nevertheless, due to the nature of the observational studies, such as is survey research, researchers do not often have control or information of the nonrespondents; thus, the study of the impact of unit nonresponse in a survey is not always doable.
- *Addressing nonresponse in surveys can be done in two ways:*

1- Up front: Designing surveys that reduces nonresponse.

2- Back-end: The use of sophisticated methods to adjust for nonresponse (e.g., Multiple Imputation and Weighting).

Transitioning to Part 2 of the Webinar

Part II: Demonstrates the impact of ignoring unit nonresponse in survey research and using complete-case analysis.

Transitioning to the Part 2 of the webinar (1/3)

Day-to-day researchers largely ignore these statistical advances because of a belief that a small amount of missing data will only slightly distort the usual estimates of means, standard deviations, correlations, and associated hypothesis tests (i.e., a continuity principle), as pointed out by Lind and Zumbo (1993).

A pernicious contributing factor is the limited availability of easy-to-use techniques in widely used statistical software programs such as SPSS and SAS. We believe, however, that the software will come along and respond to market demands when researchers insist on having statistical methods at their disposal.

Transitioning to the Part 2 of the webinar (2/3)

It should be noted that, to date, a large portion of the statistical survey research literature on unit nonresponse has focused on proposing and mathematically studying methods for adjusting for unit nonresponse rather than documenting the impact of unit nonresponse, per se.

As if the impact of unit nonresponse is widely known and appreciated, and it most certainly is for those statisticians working on the statistical theory of survey data; so, they have focused on providing and comparing solutions to what is known but, at least for applied researchers, widely undocumented statistical problem.

Transitioning to the Part 2 of the webinar (3/3)

The purpose of this webinar series is to demonstrate to day-to-day researchers the impact of unit nonresponse on conclusions from survey data and, along the way, shed some light on apparent contradictory conclusions in the literature to date.

The set of three interconnected webinars will continue as following:

Part 2 will demonstrate the impact of ignoring unit nonresponse in survey research and using complete-case analysis.

Part 3 will describe a framework for approaching the analysis of survey data (that has unit nonresponse) and demonstrate using a workable approach to adjust for unit nonresponse—propensity score weighting.

Thank you

Nonparticipation (Unit Nonresponse) In Surveys:
A Practitioner's Guide to the Conceptualization, Impact of, and Adjustment for Unit Nonresponse

Pamela Woitschach, Ph.D.

Post-Doctoral Fellow, Patient-Centred Measurement Methods
Cluster BC SUPPORT Unit
Post-Doctoral Research Fellow, UBC-Paragon Research Initiative
University of British Columbia

Bruno D. Zumbo, Ph.D.

Professor & Distinguished University Scholar
Tier 1, Canada Research Chair in Psychometrics and Measurement; &
Paragon UBC Professor of Psychometrics and Measurement
University of British Columbia

References

Beaumont, J. F. (2000). An estimation method for nonignorable nonresponse. *Survey Methodology, Statistics Canada, Catalogue No. 12-001, 26*(2), 131-136. https://www150.statcan.gc.ca/n1/pub/12-001-x/2000002/article/5532-eng.pdf

Bethlehem, J., Cobben, F., & Schouten, B. (2008). Indicators for the Representativeness of Survey Response. In Statistics Canada, *Data Collection: Challenges, Achievements and New Directions* Statistics Canada Symposium 2008,

Brick, J. M. (2013). Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics, 29*(3), 329-353. https://doi.org/10.2478/jos-2013-0026

Brick, J. M., & Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research, 5*, 215-238. https://journals.sagepub.com/doi/pdf/10.1177/096228029600500302

Carlsson, F., Merlo, J., Lindstrom, M., Ostergren, P., & Lithman, T. (2006). Representativity of a postal public health questionnaire survey in Sweden, with special reference to ethnic differences in participation. *Scandinavian Journal of Public Health, 34*, 132–139. https://doi.org/10.1080/14034940510032284

Eekhout, I., de Boer, M. R., Twisk, J. W. R., de Vet, H. C. W., & Heymans, M. W. (2012). Missing Data A Systematic Review of How They Are Reported and Handled. Epidemiology, 23(5), 729-732. https://doi.org/10.1097/EDE.0b013e3182576cdb

Chi, Y., Hong, J., Jurek, A., Liu, W., & O'Reilly, D. (2017). Privacy preserving record linkage in the presence of missing values. *Information Systems*. https://doi.org/10.1016/j.is.2017.07.001

Groves, R. M., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly, 74*(5), 849-879. https://doi.org/10.1093/poq/nfq065

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. A Meta-Analysis. *Public Opinion Quarterly, 72*(2), 167–189. https://doi.org/10.1093/poq/nfn011

Haziza, D. (2009). Imputation and inference in the presence of missing data. In D. Pfeffermann & C. R. Rao (Eds.), *Handbook of Statistics. Sample Surveys: Design, Methods and Applications*. Elsevier. https://doi.org/10.1016/S0169-7161(08)00010-2

Heeringa, S., West, B. T., & Berglund, P. (2010). *Applied Survey Data Analysis*. Chapman & HallCRC.

Kish, L. (1959). Some statistical problems in research design *American Sociological Association 24*(3), 328-338.

Koch, A., & Blohm, M. (2016). *Nonresponse bias*. GESIS-Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_004

References

Lavrakas, P. (2013). Applying a Total Error Perspective for Improving Research Quality in the Social, Behavioral, and Marketing Sciences. Presidential Address from the AAPOR 68th Annual Conference. https://www.aapor.org/About-Us/History/Presidential-Addresses/2013-Presidential-Address.aspx

Lepkowski, J. (2020). Nonresponse & Noncoverage Weighting (Coursera-University of Michigan) https://www.coursera.org/lecture/sampling-methods/6-4-nonresponse-noncoverage-weighting-4Zo0D

Little, R., & Rubin, D. B. (2002). *Statistical analysis with missing data* (Vol. 2nd.). John Wiley & Sons, Inc. .

Lind, J. C., & Zumbo, B. D. (1993). The continuity principle in psychological research: An introduction to robust statistics. Canadian Psychology, 34, 407-414.

Matei, A., & Ranalli, G. (2015). Dealing with non-ignorable nonresponse in survey sampling: A latent modeling approach. *Survey Methodology, Statistics Canada, Catalogue No. 12-001-X, 41*(1), 145-164. https://www150.statcan.gc.ca/n1/en/pub/12-001-x/12-001-x2015001-eng.pdf?st=Uq1Sc0nr

National Research Council. (2013). *Nonresponse in Social Science Surveys: A Research Agenda*. The National Academies Press. https://doi.org/10.17226/18293.

Shadish, W., Cook, T., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* Houghton Mifflin.

Statistics Canada. (2003). *Survey methods and practices*. https://www150.statcan.gc.ca/n1/en/pub/12-587-x/12-587-x2003001-eng.pdf?st=YgLx6qTD

Tourangeau, R. (2003). Cognitive aspects of survey measurement and mismeasurement. *International Journal of Public Opinion, 15*(1), 3-7. https://doi.org/10.1093/ijpor/15.1.3

Zumbo, B. D. (2007). Validity: Foundational issues and statistical methodology. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (Vol. 26 Psychometrics, pp. 45-79). Elsevier Science. https://doi.org/10.1016/S0169-7161(06)26003-6